

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2003, 15:49:18 ; Search time 48 Seconds

(without alignments)
2195.718 Million cell updates/sec

Title: US-09-590-991-6

Perfect score: 4398
Sequence: 1 YPVVLAATSSSEDALNISDK.....PGKNIKSTLVINGKSTGY 664

Scoring table: BIOSUM30
Gapop 8.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4398	100.0	664	22	AA848342
2	4398	100.0	2120	21	AA848342
3	4380.5	99.6	2140	24	ABU01020
4	4171	94.8	1007	19	AAW61246
5	4171	94.8	1007	23	ABP54664
6	975.5	22.2	192	19	AA186159
7	962.5	21.9	1962	12	AA10559
8	954.5	21.7	1968	12	AA10941
9	953.5	21.7	1962	12	AA10557

10	953.5	21.7	1974	12	AA10940
11	951.5	21.6	1962	12	AA10558
12	948.5	21.6	1962	12	AA10560
13	948.5	21.6	1962	12	AA10561
14	948.5	21.6	1962	12	AA10563
15	947	21.5	1959	12	AA10562
16	938.5	21.3	1560	10	AA10562
17	938.5	20.4	1946	19	AAW47273
18	896.5	19.9	1570	23	ABP29894
19	875	19.9	1570	23	ABP29894
20	875	19.9	1570	23	ABP29894
21	875	19.9	1570	23	ABP29894
22	869.5	19.8	1550	23	ABP29894
23	868	19.7	885	23	ABP29796
24	867.5	19.7	1239	23	ABP25832
25	865.5	19.7	1579	24	ABP71253
26	864	19.6	1233	23	ABP30203
27	862	19.6	1233	23	ABP29675
28	857	19.5	1647	23	ABP28561
29	797.5	18.1	690	11	AA104904
30	792	18.0	2710	17	AA105016
31	792	18.0	2710	17	AA105016
32	787.5	17.9	690	18	AAW16337
33	787.5	17.9	690	18	AAW16337
34	782	17.8	806	13	AA107481
35	780.5	17.7	2835	23	AB198574
36	774	17.6	815	23	AB198574
37	771	17.5	922	19	AAW37372
38	769	17.5	1475	11	AA108221
39	764.5	17.4	1582	23	AB107307
40	761.5	17.3	903	17	AA107007
41	761.5	17.3	1398	17	AA107008
42	761.5	17.3	1398	18	AAW4124
43	761.5	17.3	1398	20	AAW4839
44	761	17.3	1941	23	ABP5509
45	760.5	17.3	1185	13	AA122675

ALIGNMENTS

RESULT 1
ID AAB48342 standard; Protein: 664 AA.
AC AAB48342;
DT 20-APR-2001 (first entry)
XX S. pneumoniae Spi28 polypeptide.
XX Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;
XX bronchial; lung; blood; infection; immune response; immunotherapy;
XX antibacterial; auditory; vaccine.
XX Streptococcus pneumoniae.
XX WO200076540-A2.
XX 21-DEC-2000.
XX PD 09-JUN-2000; 2000WO-US15925.
XX PF 10-JUN-1999; 99US-0138453.
XX PR (MEDI-) MED IMMUNE INC.
XX PA Adamou JE, Choi GH;
XX WI; 2001-112197/12.
XX DR N-PSDB; AAC84741.
XX PT New vaccines comprising Spi28 or Spi30 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
XX blood infections -
XX
XX
XX Claim 4; Page 47-50; 54pp; English.
XX
CC The invention relates to novel immunogenic polypeptides, Spl28 and Spl30
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC the treatment and prevention of pneumococcal infections, particularly
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC bronchial, lung or blood infections. The antigens are used as immunogenic
CC agents to stimulate an immune response. The antisera and antibodies may
CC also be used in diagnosing and treating pneumococcal infections.
CC Recombinant polypeptides serve as a mechanism for stimulating production
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC as reagents in other processes such as affinity chromatography. The
XX present sequence represents the S. pneumoniae Spl28 polypeptide.
XX
SQ Sequence 664 AA;

Query Match 100.0%; Score 4398; DB 22; Length 664;
Best Local Similarity 100.0%; Pred. No. 4e-64;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPVVLADTSSSEDALNISKKEVAENKEKHENIHSAMETSDQFKEKKTAVIKEKEVVSKN 60
DB 1 YPVVLADTSSSEDALNISKKEVAENKEKHENIHSAMETSDQFKEKKTAVIKEKEVVSKN 60

QY 61 PVIDNNTSNEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120
DB 61 PVIDNNTSNEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120

QY 121 KELSSLKNTKVLTYDRIFNGSAIETTPDNLDKIKQIEGSISSVERAQKVPMMNHARKEI 180
DB 121 KELSSLKNTKVLTYDRIFNGSAIETTPDNLDKIKQIEGSISSVERAQKVPMMNHARKEI 180

QY 181 GVEAIDYLSINAPFGKNGFGRGMVINSIDTGDYRHKAMRIDDDAKASMRFKKEDLKG 240
DB 181 GVEAIDYLSINAPFGKNGFGRGMVINSIDTGDYRHKAMRIDDDAKASMRFKKEDLKG 240

QY 241 TDKNYWLSDKIPIAFNYNGGKITVEKYDDGRDYFDPHGHIAGILAGNDTEQDIKNFNG 300
DB 241 TDKNYWLSDKIPIAFNYNGGKITVEKYDDGRDYFDPHGHIAGILAGNDTEQDIKNFNG 300

QY 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVVDVSVSSGGFTGTGLVGEKY 360
DB 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVVDVSVSSGGFTGTGLVGEKY 360

QY 361 WQAIRLRKAGIPMVATGNVYATSSSSWDLVANNHLKMTDTCGNVTRTAHEDAIAVAS 420
DB 361 WQAIRLRKAGIPMVATGNVYATSSSSWDLVANNHLKMTDTCGNVTRTAHEDAIAVAS 420

QY 421 AKNOTVEFDKVNIGGESFKYRNIGAFDPSKITTNEDGTAKPSKLKFVYIGKQDQDLIG 480
DB 421 AKNOTVEFDKVNIGGESFKYRNIGAFDPSKITTNEDGTAKPSKLKFVYIGKQDQDLIG 480

QY 481 LDLRGKIADVNDRIYTKDLKNAFKKAMDKGARAIMVNTVNYNRDNTMELPAMGYEADG 540
DB 481 LDLRGKIADVNDRIYTKDLKNAFKKAMDKGARAIMVNTVNYNRDNTMELPAMGYEADG 540

QY 541 TKSOVFSISGDDGKVLNMTNPDKKTVEKRNKEDPKDLQYYPIDMESFNSKNPNVGD 600
DB 541 TKSOVFSISGDDGKVLNMTNPDKKTVEKRNKEDPKDLQYYPIDMESFNSKNPNVGD 600

QY 601 EKEIDFKFAPDQDKELYKEDIIVPAGTSWGPRIIDLLKPDVSPAGPKNIKSTLNVINGS 660
DB 601 EKEIDFKFAPDQDKELYKEDIIVPAGTSWGPRIIDLLKPDVSPAGPKNIKSTLNVINGS 660

QY 661 TYGY 664
DB 661 TYGY 664

RESULT 2
AAY81710
ID AAY81710 standard; Protein; 2120 AA.
XX
XX AAY81710;
AC
XX
DT 02-JUN-2000 (first entry)
XX
DE Streptococcus pneumoniae protein sequence ID3.
XX
XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
XX Streptococcus pneumoniae.
OS
XX WO200006738-A2.
PN
XX
XX 10-FEB-2000.
PD
XX 27-JUL-1999; 99WO-GB02452.
PF
XX 19-MAR-1999; 98GB-0016336.
PR
XX 27-JUL-1998; 99US-0125329.
PR
XX (MICR-) MICROBIAL TECHNIQS LTD.
PA
XX
XX
PI Le Page RWF, Wells JM, Hanniffy SB, Hanabro PM;
PI WPI; 2000-195301/17.
DR N-PSDB; AAZ91806.
DR
XX
XX Streptococcal proteins and polynucleotides useful for diagnosis,
PT treatment and prophylaxis of bacterial infections -
XX
XX Claim 2; Page 41-42; 76pp; English.
PS
XX This sequence represents a Streptococcus pneumoniae protein of the
CC invention. The proteins (or their homologues, derivatives and/or
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
CC compositions comprising the proteins are useful as vaccines and also in
CC diagnostic assays. The sequences are useful for the detection or
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
CC with them. Agents capable of antagonising, inhibiting or interfering with
CC the function or expression of the protein or polypeptide are useful in
CC medical compositions in the treatment or prophylaxis of S. pneumoniae
CC infection. As the sequences can be used to treat S. pneumoniae infection,
CC they can be used to treat bacterial pneumonia, which has high rates in
CC young children, the elderly, and in patients with predisposing conditions
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC or with immunosuppressive disorders, especially AIDS. They can also be
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC meningitis.
XX
SQ Sequence 2120 AA;

Query Match 100.0%; Score 4398; DB 21; Length 2120;
Best Local Similarity 100.0%; Pred. No. 9.6e-63;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPVVLADTSSSEDALNISKKEVAENKEKHENIHSAMETSDQFKEKKTAVIKEKEVVSKN 60
DB 1 YPVVLADTSSSEDALNISKKEVAENKEKHENIHSAMETSDQFKEKKTAVIKEKEVVSKN 60

QY 61 PVIDNNTSNEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120
DB 61 PVIDNNTSNEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120

QY 121 KELSSLKNTKVLTYDRIFNGSAIETTPDNLDKIKQIEGSISSVERAQKVPMMNHARKEI 180
DB 121 KELSSLKNTKVLTYDRIFNGSAIETTPDNLDKIKQIEGSISSVERAQKVPMMNHARKEI 180

QY 181 GVEEAIDYLSKINAFPGNFDGRGVISNIDTGTDIRRHKARIDDDAKASMRFKKEDLKG 240
DB 181 GVEEAIDYLSKINAFPGNFDGRGVISNIDTGTDIRRHKARIDDDAKASMRFKKEDLKG 240
QY 241 TDKRYWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHAGIAGLNDTEODIKPNFG 300
DB 241 TDKRYWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHAGIAGLNDTEODIKPNFG 300
QY 301 IDGIAFPAQAQISYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
DB 301 IDGIAFPAQAQISYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
QY 361 WQAIRALARKAGIPMVVATGNVATSASSSSWDLVANNHJKMTDGTGNVTRTAHEDAIAVAS 420
DB 361 WQAIRALARKAGIPMVVATGNVATSASSSSWDLVANNHJKMTDGTGNVTRTAHEDAIAVAS 420
QY 421 AKNOTVEFDKVNIGESFKYRNIGAFPDKSKITTNEDGTAPSKLKFYIGKGDODLIG 480
DB 421 AKNOTVEFDKVNIGESFKYRNIGAFPDKSKITTNEDGTAPSKLKFYIGKGDODLIG 480
QY 481 LDLRGKIAVMBRITYTKDLKNAFKKAMDGARAIVVNVVYVNRDNTLPMAGYEADG 540
DB 481 LDLRGKIAVMBRITYTKDLKNAFKKAMDGARAIVVNVVYVNRDNTLPMAGYEADG 540
QY 541 TKSQVFSISGDDGVKLMNNINPDCKTEYKRNKEDFKDLKQYYPIDMESFNSNKPNGD 600
DB 541 TKSQVFSISGDDGVKLMNNINPDCKTEYKRNKEDFKDLKQYYPIDMESFNSNKPNGD 600
QY 601 EKEIDFKFAPDPTDKELYEDIIIVPAGSTSWGPRIDLKLPVSAFGKNIKSTLVNINGS 660
DB 601 EKEIDFKFAPDPTDKELYEDIIIVPAGSTSWGPRIDLKLPVSAFGKNIKSTLVNINGS 660
QY 661 TYGY 664
DB 661 TYGY 664

RESULT 3
ABU01020
ID ABU01020 standard; Protein; 2140 AA.

XX AC ABU01020;
XX DT 11-FEB-2003 (first entry)
XX DE S. pneumoniae type 4 strain protein from coding region #590.
XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;
XX KM ear infection; antiinflammatory; antibacterial; immunostimulant;
XX KM auditory; respiratory; gene therapy; vaccine.
XX OS Streptococcus pneumoniae type 4 strain.
XX PN WQ200277021-A2.
XX PD 03-OCT-2002.
XX PF 27-MAR-2002; 2002WO-IB02163.
XX PR 27-MAR-2001; 2001GB-0007658.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Maelignant V, Tectelin H, Fraser C;
XX DR WPI; 2003-040579/03.
XX DR N-PSDB; ABX06302.
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection

XX XX Claim 1; SEQ ID No 1180; 56bp; English.
PS The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2469 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the terminus of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The method are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX XX Sequence 2140 AA;

Query Match 99.6%; Score 4380.5; DB 24; Length 2140;
Best Local Similarity 99.6%; Pred. No. 1.9e-67;
Matches 664; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

QY 1 Y-P--VVLADTSSSEDLNLSDEKEVAENKEKHENISAMETSODFEKKTAVIKEKEV 57
DB 19 YAPNEVLAOTRSSSEDLNLSDEKEVAENKEKHENISAMETSODFEKKTAVIKEKEV 78
QY 58 SKNPVINDNTSNEBAKIKEENSNGSGDYTDSFVNKNTENPKKEDKYVYIAEFKDSGE 117
DB 79 SKNPVINDNTSNEBAKIKEENSNGSGDYTDSFVNKNTENPKKEDKYVYIAEFKDSGE 138
QY 118 KATKELSSKTKTKLYYDRIFNGSAIETTPDNLDKIKOIEGISVERAOKVQPMNHAR 177
DB 139 KATKELSSKTKTKLYYDRIFNGSAIETTPDNLDKIKOIEGISVERAOKVQPMNHAR 198
QY 178 KEIGVEEAIDYLSKINAFPGNFDGRGVISNIDTGTDIRRHKARIDDDAKASMRFKKED 237
DB 199 KEIGVEEAIDYLSKINAFPGNFDGRGVISNIDTGTDIRRHKARIDDDAKASMRFKKED 258
QY 238 LKGTDKXWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHAGIAGLNDTEODIKN 297
DB 259 LKGTDKXWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHAGIAGLNDTEODIKN 318
QY 298 FNGIDGIAFPAQAQISYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVG 357
DB 319 FNGIDGIAFPAQAQISYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVG 378
QY 358 EKYWQAIRALARKAGIPMVVATGNVATSASSSSWDLVANNHJKMTDGTGNVTRTAHEDAIA 417
DB 379 EKYWQAIRALARKAGIPMVVATGNVATSASSSSWDLVANNHJKMTDGTGNVTRTAHEDAIA 438
QY 418 VASAKNOTVEFDKVNIGESFKYRNIGAFPDKSKITTNEDGTAPSKLKFYIGKGDOD 477
DB 439 VASAKNOTVEFDKVNIGESFKYRNIGAFPDKSKITTNEDGTAPSKLKFYIGKGDOD 498

QY 478 LIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNTVYNNRDNWTLPAMGYEA 537
 DB 499 LIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNTVYNNRDNWTLPAMGYEA 558
 QY 538 DEGTQSQVFSISGDDGVKLNWNNINPDKTEVRNNKEDFKLEQYYPIDMESFNSNKP 597
 DB 559 DEGTQSQVFSISGDDGVKLNWNNINPDKTEVRNNKEDFKLEQYYPIDMESFNSNKP 618
 QY 598 VGDEKEIDFKAPDPTDKELYKEDIIVPAGSTSWGPRIDLLKPDVSAFGKNIKSTLNVIN 657
 DB 619 VGDEKEIDFKAPDPTDKELYKEDIIVPAGSTSWGPRIDLLKPDVSAFGKNIKSTLNVIN 678
 QY 658 GKSTYGY 664
 DB 679 GKSTYGY 685

RESULT 4
 AAW61246
 ID AAW61246 standard; Protein; 1007 AA.
 XX
 AC AAW61246;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae SPI22 protein.
 XX
 KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
 KW detection; pneumonia; otitis media; meningitis.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9818930-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19422.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
 XX
 DR WPI; 1998-272224/24.
 DR N-PSDB; AAV27431.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
 PT pneumoniae - or their epitope-containing fragments, useful in
 PT protective or therapeutic vaccines, and for diagnosis
 XX
 PS Claim 11; Page 92-93; 118pp; English.
 XX
 CC The present sequence represents a protein from Streptococcus pneumoniae.
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
 CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose.

XX Sequence 1007 AA;
 Query Match 94.8%; Score 4171; DB 19; Length 1007;
 Best Local Similarity 100.0%; Pred. No. 8e-60;
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ETSQDFEKKTAVTKEKEVSKNPVIDNNTSNEBAKIKEENSKSQDYTTDSFVNKNTEN 97
 DB 1 ETSQDFEKKTAVTKEKEVSKNPVIDNNTSNEBAKIKEENSKSQDYTTDSFVNKNTEN 60
 QY 98 PKCEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLTYDRIFNGSAIETTTDNLDKIKOI 157
 DB 61 PKCEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLTYDRIFNGSAIETTTDNLDKIKOI 120
 QY 158 EGISSEVRAQKVQPMNHARKEIGVEEADYKLSINAPFGKNFQGRGMVINSIDTGTDIR 217
 DB 121 EGISSEVRAQKVQPMNHARKEIGVEEADYKLSINAPFGKNFQGRGMVINSIDTGTDIR 180
 QY 218 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYNGGKITVEKYDDGDRYDPP 277
 DB 181 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYNGGKITVEKYDDGDRYDPP 240
 QY 278 HGMHIAGILAGNTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 337
 DB 241 HGMHIAGILAGNTEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 300
 QY 338 KHNVDVSVSGFTGTGLVGEKYWOAIRALRKAGIPMVVATGNVATSSSSSSDVLVANNH 397
 DB 301 KHNVDVSVSGFTGTGLVGEKYWOAIRALRKAGIPMVVATGNVATSSSSSSDVLVANNH 360
 QY 398 LKMTDTGNVTRTAHEDAIASAKNOTVEFDKYNIGGESFKYRNIGAFFDKSKITTTNED 457
 DB 361 LKMTDTGNVTRTAHEDAIASAKNOTVEFDKYNIGGESFKYRNIGAFFDKSKITTTNED 420
 QY 458 GTKAPSKLKFVYIGKQDQDLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVYN 517
 DB 421 GTKAPSKLKFVYIGKQDQDLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVYN 480
 QY 518 TVNYYNRDNWTLPAMGYEADGKTSQVFSISGDDGVKLNWNNINPDKTEVRNNKEDPK 577
 DB 481 TVNYYNRDNWTLPAMGYEADGKTSQVFSISGDDGVKLNWNNINPDKTEVRNNKEDPK 540
 QY 578 DKLEQYYPIDMESFNSNKPNGDEKEIDFKAPDPTDKELYKEDIIVPAGSTSWGPRIDLL 637
 DB 541 DKLEQYYPIDMESFNSNKPNGDEKEIDFKAPDPTDKELYKEDIIVPAGSTSWGPRIDLL 600
 QY 638 LKPDVSAFGKNIKSTLNVIKSTYGY 664
 DB 601 LKPDVSAFGKNIKSTLNVIKSTYGY 627

RESULT 5
 ABP54664
 ID ABP54664 standard; Protein; 1007 AA.
 XX
 AC ABP54664;
 XX
 DT 04-SEP-2002 (first entry)
 XX
 DE S. pneumoniae SPI22 protein sequence SEQ ID NO:216.
 XX
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN US2002061545-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 22-JAN-2001; 2001US-0765272.
 XX
 PR 30-OCT-1997; 97US-0961083.
 XX
 PA (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX
 PI Choi GH, Kunech CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX
 DR WPI: 2002-479261/51.
 DR N-PSDB; ABQ84899.
 XX
 PT New Streptococcus pneumoniae antigens, useful for detecting
 PT Streptococcus and for preventing or attenuating disease caused by
 PT Streptococcus infection -
 XX
 PS Claim 11; Page 50; 70pp; English.
 XX
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
 CC which are used in an example from the present invention.
 CC
 XX
 XX Sequence 1007 AA;
 SO
 Query Match 94.8%; Score 4171; DB 23; Length 1007;
 Best Local Similarity 100.0%; Pred. No. 8e-60;
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 ETSODFEKKTAVIKKEKEVSNKPIIDNNTSNEAKTEENSOGGYTDSFVNKNTEN 97
 DB 1 ETSODFEKKTAVIKKEKEVSNKPIIDNNTSNEAKTEENSOGGYTDSFVNKNTEN 60
 QY 98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTDRIFFNGSAIETTPDNLKIKOI 157
 DB 61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTDRIFFNGSAIETTPDNLKIKOI 120
 QY 158 EGISSEVERAQQVQPMNHARKEIGVEAIDVYKLSINAPFGKNPDGRGVNISNIDTGYR 217
 DB 121 EGISSEVERAQQVQPMNHARKEIGVEAIDVYKLSINAPFGKNPDGRGVNISNIDTGYR 180
 QY 218 HKARIDDDAKASMFKEKEDLKGTDKNWLSDKIPIHANNYNGKITTEKXDGRDYDP 277
 DB 181 HKARIDDDAKASMFKEKEDLKGTDKNWLSDKIPIHANNYNGKITTEKXDGRDYDP 240
 QY 278 HGMHAGILAGNDTEODIKFNFGIDGIAIPNAQIFSYKNYSDAGSGFAGDETFHAIEDSI 337
 DB 241 HGMHAGILAGNDTEODIKFNFGIDGIAIPNAQIFSYKNYSDAGSGFAGDETFHAIEDSI 300
 QY 338 KHNVVVSVSSGFTGTGLVGEKXWQAIRALKRAGIPMYVATGVNYATSSASSSMDLVANNH 397
 DB 301 KHNVVVSVSSGFTGTGLVGEKXWQAIRALKRAGIPMYVATGVNYATSSASSSMDLVANNH 360
 QY 398 LKMTDTGVNTTAHEDIAAASAKNQVPEPKNVIIGSESPRYRITGAFPPKSKITTTMED 457
 DB 361 LKMTDTGVNTTAHEDIAAASAKNQVPEPKNVIIGSESPRYRITGAFPPKSKITTTMED 420
 QY 458 GTKAPSKLKFVYIGGQODDLIGDLRGKIAVMRIYTKDLKNAFKKMDKARAIMVYN 517
 DB 421 GTKAPSKLKFVYIGGQODDLIGDLRGKIAVMRIYTKDLKNAFKKMDKARAIMVYN 480
 QY 518 TVNYNNDNMTLPMWGYEADGTSQVFSISGDDGVKLMNINDPDKTEVRANKKEDPK 577
 DB 481 TVNYNNDNMTLPMWGYEADGTSQVFSISGDDGVKLMNINDPDKTEVRANKKEDPK 540
 QY 578 DKLQGYVIDMESFNSKNPNVGEDEKIDFKPAPDIDKLYKEDIIVPAGSTMGWRIDL 637
 DB 541 DKLQGYVIDMESFNSKNPNVGEDEKIDFKPAPDIDKLYKEDIIVPAGSTMGWRIDL 600
 QY 638 LKPDVSAFGKNIKSTLVINGKSTYGY 664

DB 601 LKPDVSAFGKNIKSTLVINGKSTYGY 627
 |||
 RESULT 6
 ID AAY86159 standard; Protein; 192 AA.
 XX
 AC AAY86159;
 XX
 DT 10-APR-2000 (first entry)
 XX
 DE S. pneumoniae derived protein #368.
 XX
 KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN MO9806734-A1.
 XX
 PD 19-FEB-1998.
 XX
 PF 15-AUG-1997; 97WO-US14436.
 XX
 PR 16-AUG-1996; 96US-0024022.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Stodola RK;
 XX
 DR WPI: 1998-159452/14.
 DR N-PSDB; AAZ96473.
 XX
 PT Streptococcus pneumoniae proteins and related DNA - useful for
 PT screening compounds for antibacterial activity
 XX
 PS Claim 5; Page 617; 640pp; English.
 XX
 CC This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see AAZ96173-296494) and their encoded proteins (see
 CC AAZ85792-Y86182). The DNA, vectors and host cells described in the
 CC method of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.
 CC
 XX
 SO Sequence 192 AA;
 Query Match 22.2%; Score 975.5; DB 19; Length 192;
 Best Local Similarity 93.7%; Pred. No. 3.3e-08;
 Matches 163; Conservative 1; Mismatches 7; Indels 3; Gaps 2;
 QY 1 Y-P--VVLADSSSEDALNISDKKVAENKEKHINHSAMETSDPFKKTAVIKKEKVV 57
 DB 19 YAPNEVVLADTSNEDALNISDKKVAENKEKHINHSAMETSDPFKKTAVIKKEKVV 78
 QY 58 SKNPVINNTSNEBAKIKENSNSKQGDYTDSPFNKNTENPKKEDKVVYIAEFKDKESGE 117
 DB 79 SKNPVINNTSNEBAKIKGGSNSKQGDYTDSPFNKNTENPKKEDKVVYIAEFKDKESGE 138
 QY 118 KAIKELSSLKNTKVLYTDRIFFNGSAIETTPDNLKIKQIEGISSVERAQQVOP 171
 DB 139 KAIKELSSLKNTKVLYTDRIFFNGSAIETTPDNLKIKQIEGISSVERAQQVHP 192

RESULT 7	
ID	AAR10559 standard; Protein; 1962 AA.
XX	
AC	AAR10559;
XX	
DT	15-APR-1991 (first entry)
XX	
DE	Mutant protease (A137G/K138L/T139A).
XX	
KW	Mutant protease gene; fermentation; foodstuff; flavouring;
KW	lactic acid bacteria.
XX	
OS	Lactococcus lactis SK11.
XX	
PH	Key
FT	Location/Qualifiers
FT	1..187
FT	/label= signal_peptide
FT	Protein
FT	188..1962
FT	/label= mature_protein
XX	
PN	EP411715-A.
XX	
PD	06-FEB-1991.
XX	
PF	02-AUG-1990; 90EP-0202113.
XX	
PR	04-AUG-1989; 89NL-0002010.
XX	
PA	(NEZU-) NED INST ZUIVELONDE.
XX	
PI	Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX	
DR	WPI; 1991-038622/06.
DR	N-PSDB; AAQ10413.
XX	
PT	Mutant protease gene(s) and protease(s) - derived from type I and
PT	III protease genes from lactococcal strains, used in fermentation
PT	foodstuffs and flavourings
XX	
PS	Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX	
CC	The mutant protease having new cleavage specificities is
CC	obtained by replacing three amino acids.
CC	This mutant may then be used to prepare hybrid proteases,
CC	the fusion being between a type I and a type III protease of
CC	L.lactis Wg2 and SK11.
CC	The product has modified properties, e.g. thermostability,
CC	alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC	compared to the parent protease(s). The proteases can be used for
CC	preparing products (butter cheese, human and animal foodstuffs)
CC	prepared with the aid of lactic acid bacteria.
CC	See also AAQ10411-17 and AAQ10870-71.
XX	
SQ	Sequence 1962 AA;
	Query Match 21.9%; Score 962.5; DB 12; Length 1962;
	Best Local Similarity 26.1%; Pred. No. 3.2e-05;
	Matches 182; Conservative 134; Mismatches 241; Indels 139; Gaps 33;
QY	1 YPV--VLADTSSEBALNISDEKVAENKEKHENHSAHMETSQDPFKEKTAIVIKEVVS 58
Db	24 LPVGEIOAKAASQOTLSSANTVTAA-----TAKQAATD-----TTA-----AT 64
QY	59 KNPVINDNNTSNEAKIKENSNGDGYTDSFVNKNTENPKKEDKVYIAEPKDKESGK 118
Db	65 TNQATATQLAAGIDYNKLVKQVQDDIYVDVIVQM--SAAPASENGILRTDYSSIAEQE 123
QY	119 AIKELSSLNKNTKLVY-----TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAOK 168
Db	124 TNKVIAQAASVKAAVEQVTTQTAGESGYVVGFGSTKVRVWDIPKLQIAGVKTIVLAKV 183
QY	169 VQPMNHARKEIGVEEA-IDYLSINAPFGKFNDRGMVSNIDTGTDRHKAMRIDDDA 227
Db	184 YPTDAKANSMANVQAVMSNY-----KYKGEIVSVVIDSGIDPTHKDMRLSDDK 233
QY	228 KASMRFKKEDL-KGTDK---NYWLSDKTIPAFNY-YNGGKITVEKYDGRDYDFPHGWHI 282
Db	234 DV--KLTKSDVEKFTDTVKHGRYFNSKVPYGFVADNNDTIT---DDKVD--EQHGMHV 285
QY	283 AGILAGNTEQDIKNFNGIDGIAPNAQIFSKYKMSYSDAG-SGFAGDETWFHAIEDSIKHN 341
Db	286 AGIIGANGTGDPA--KSVVGVAPPAQLAMKVFNSDTSGLAGSATVVSIAEDSAKIGA 343
QY	342 DVVSVSSG-FTGTGLVGEKYQOAIRALKRAGIPMVVATGNVATSSSSSWDLVANNHLM 400
Db	344 DVLNMSLGSNGNQTLEDPAAVQVONANESGTAAVISAGNSCTSGSATEG--VNKDYVGL 401
QY	401 TDTGNVTRTAAHEDAIAVASAKNOTVEFDKVI-----GGESFKYRN---IGAFDDK 449
Db	402 QDNEMVSGSPGTSRGATTVASAENTDVIQAVTITDGTGLQLGPETIQLSSHDFTGS-PDQ 460
QY	450 SKITTNEGTAKPSKLFVYIGKGQDQLIGLDLKGIAVMDRI-YTKDLKNVAFKAMDK 508
Db	461 KKFYIVKDAAGN-----LSKALADYTA-DAKGKIAIVKRGESFDDKQKQAQA-- 509
QY	509 GARAIMVNTVNYNRDNWTELPAWGYEADGCTKSQVFSISGDDGVKL--WNMINPKKT 566
Db	510 GAAGLIIVNT-----DGTATPMT--SIALTTTPTFGLSSVTQKLVDTWTAHPDDSL 560
QY	567 EYKRNNKEDFKDLE-QYYPIDMESFNSKNPNVGDEKIDFKFAPDTPDKELYKEDIIVPA 625
Db	561 GVKIT-----LAMLPNQYTEDKMS-----DF----- 582
QY	626 GSTSHGPRIDILLKPDVSAPGKNIKSTLNVINGKST 661
Db	583 --TSYGPVSNLSFKPDITAPGGINWSTQN--NNGYTN 615
RESULT 8	
ID	AAR10941 standard; Protein; 1968 AA.
XX	
AC	AAR10941;
XX	
DT	15-APR-1991 (first entry)
XX	
DE	Mutant protease (delta137-139/ins9).
XX	
KW	Mutant protease gene; fermentation; foodstuff; flavouring;
KW	lactic acid bacteria.
XX	
OS	Lactococcus lactis SK11.
XX	
PH	Key
FT	Location/Qualifiers
FT	1..187
FT	/label= sig_peptide
FT	Protein
FT	188..1968
FT	/label= mat_protein
XX	
PN	EP411715-A.
XX	
PD	06-FEB-1991.
XX	
PF	02-AUG-1990; 90EP-0202113.
XX	
PR	04-AUG-1989; 89NL-0002010.
XX	
PA	(NEZU-) NED INST ZUIVELONDE.
XX	
PI	Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX	
DR	WPI; 1991-038622/06.
DR	N-PSDB; AAQ10871.
XX	


```

Db 124 TNKVIAQAASVKAAVEQVQTQTAGESYGVVNGFSTKVRVVDIPKLGKQIAGVKTVTAKV 183
Qy 169 VQPMNHARKEIGVEEA-IDYLSINAPFGKNGFGRGMVINSIDTGTDIRHAKMRIDDDA 227
Db 184 YPTDANKANSMAVQAVMSNY-----KYKGEFTVWSVIDSGIDPTHKMRSLSDK 233
Qy 228 KASMRFKKEDL-KGTDK---NYWLSDKTIPHAFNY-YNGGKITVEKYDGRDYFDPHGMHI 282
Db 234 DV--KLTKSDVEKFTDTVHKGRYFNSKVPYGFNYADNDTIT---DDKVD--EQHGMHV 285
Qy 283 AGILAGNDTEODIKNFNGIDGIAPNAQIFSYKMYSDAG-SGFAGDETTFHAIEDSIKENV 341
Db 286 AGIIGANGTGDDPA--KSVVGVAPPAQLAMKVFNSDTSKGTSGATVVSIAEDSAKIGA 343
Qy 342 DVSVSSG-FTGTGLVGEKYQWAIIRALKAGIPMVVATGNVATSASSSSDLVANNHLM 400
Db 344 DVLNMSLGSNSGNTLEDELAQVNAESGTAAVISAGNSGTSGATEG--VNKDYGYL 401
Qy 401 TDGNTVTRTAHEDAIAVASAKNOTVEFDKNI-----GGESFKYRN---ICAFEDK 449
Db 402 QDNEMVSGPSTSGATTVASAENTDVIQAVTITDGTGLQGPETIQUSHDFTGS-EDQ 460
Qy 450 SKITTNEDGTPKAPSKLVYTGKQDQDLGLRGLKIAVMDRI-YTKDLKNAFKKAMD 508
Db 461 KKFVIVKDSGN-----LSKGLADYTA-DAKGKIAIVKRGESFDDKQYQAQA-- 509
Qy 509 GARAIMVNTVNYNRDNTWTELPAMGYEADSGTKSQVFSISGDDGVKL--WNMINPKKT 566
Db 510 GAAGLIIVNT-----DGTATPMT--SIALTTTFFTEGLSVTQKLVDMVTAHPDDSL 560
Qy 567 EVKRNKEDFKDLE-QYYPIDMESFNSKNPVGDEKEIDKFAPDTKELYKEDIIVPA 625
Db 561 GVKIT-----LAMPLNPKYTEDKMS-----DF----- 582
Qy 626 GSTSGWPRIDLLKPDVSAPGKNTKSTLNVLNGKST 661
Db 583 --TSYGPVSNLSFPRDITAPGNIWSTQN--NNGYTN 615

RESULT 10
AAR10940
ID AAR10940 standard; Protein; 1974 AA.
XX
AC AAR10940;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (delta137-139/ins15).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT Protein /label= SIG PEPTIDE
FT 188..1974
FT /label= MAT_PROTEIN
XX
PN EP411715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI; 1991-038622/06.

```

```

DR N-PSDB; AAQ10870.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(b); 29pp ; English.
XX
CC The wild-type L.lactis SK11 protease gene sequence was determined
CC by the applicant (EP-307011).
CC The mutant protease having new cleavage specificities is
CC obtained by deleting three amino acids (nine bps) and inserting
CC 15 other residues.
CC This mutant may then be used to prepare hybrid proteases, of
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1974 AA;
Query Match 21.7%; Score 952.5; DB 12; Length 1974;
Best Local Similarity 25.8%; Pred. No. 4.8e-05;
Matches 183; Conservative 133; Mismatches 241; Indels 151; Gaps 35;
Qy 1 YPV--VLADTSSSDALNISDKKVAENKEHENIHSAMETSQDPKPKTAIVKEKVVVS 58
Db 24 LPVGEIQAKAISQOTLGSSLANVTAA-----TAKQAATD-----TTA-----AT 64
Qy 59 KNPVIDNNTSNEEAKIKEENSKSQGYDTSFVNKNTENPKKEDKVVIAEPKDESGBK 118
Db 65 TNQAIATQLAAKGIDYNKLNKVVQODIYVDVIVQM-SAAPASENGILRTDYSSTAIEIQE 123
Qy 119 AIKELSSLKNTKLY-----TYDRIFNGSAIETTPDNLKIKOIEGISSVERAQK 168
Db 124 TNKVIAQAASVKAAVEQVQTQTAGESYGVVNGFSTKVRVVDIPKLGKQIAGVKTVTAKV 183
Qy 169 VQPMNHARKEIGVEEA-IDYLSINAPFGKNGFGRGMVINSIDTGTDIRHAKMRIDDDA 227
Db 184 YPTDANKANSMAVQAVMSNY-----KYKGEFTVWSVIDSGIDPTHKMRSLSDK 233
Qy 228 KASMRFKKEDL-KGTDK---NYWLSDKTIPHAFNY-YNGGKITVEKYDGRDYFDPHGMHI 282
Db 234 DV--KLTKSDVEKFTDTVHKGRYFNSKVPYGFNYADNDTIT---DDKVD--EQHGMHV 285
Qy 283 AGILAGNDTEODIKNFNGIDGIAPNAQIFSYKMY--SD-----AG-----SGFAGDET 329
Db 286 AGIIGANGTGDDPA--KSVVGVAPPAQLAMKVFNSDTSKGTGLAGTKTGKGTGSATV 343
Qy 330 FHAIEDSIKHNVDVSVSSG-FTGTGLVGEKYQWAIIRALKAGIPMVVATGNVATSASSS 388
Db 344 VSAIEDSAKIGADVNLNMSLGSNSGNTLEDELAQVNAESGTAAVISAGNSGTSGSAT 403
Qy 389 SWDLVANNHLMKMTDGNVTRTAHEDAIAVASAKNOTVEFDKNI-----GGESFKY 440
Db 404 EG--VNKDYGYGLQDNEMVSGPSTSGATTVASAENTDVIQAVTITDGTGLQGPETIQ 461
Qy 441 RN---IGAFFDKSKITTNEDGTPKAPSKLVYTGKQDQDLGLRGLKIAVMDRI-YTK 496
Db 462 SSHDFTGS-FDQKFPYIVKDSGN-----LSKGLADYTA-DAKGKIAIVKRGESF 511
Qy 497 DLKNAFKKAMDKGARAIMVNTVNYNRDNTWTELPAMGYEADSGTKSQVFSISGDDGVKL 556
Db 512 DDQKQYQAQA--GAAGLIIVNT-----DGTATPMT--SIALTTTFFTEGLSVTQKLV 560
Qy 557 --WNMINPKKTEVKRNKEDFKDLE-QYYPIDMESFNSKNPVGDEKEIDKFAPD 613
Db 561 VDMVTAHPDDSLGVKIT-----LAMPLNPKYTEDKMS-----DF----- 594

```


Qy 614 KELYKEDIIVAGSTSGPRIDLLKPDVSAPGKNIKSTLVNIGKST 661
 Db 595 -----TSYGPVSNLSFKPDITAPGGINIMSTON--NNGYTN 627

RESULT 11

AA010558
 ID AAR10558 standard; Protein; 1962 AA.

XX AAR10558;

XX 15-APR-1991 (first entry)

XX Mutant protease (A137G/K138P/T139P).

XX Mutant protease gene; fermentation; foodstuff; flavouring;

XX Lactic acid bacteria.

XX Lactococcus lactis SK11.

XX Key Location/Qualifiers

FT Peptide 1..187

FT Protein /label= signal_peptide

FT Protein /label= mature_protein

XX EP411715-A.

XX 06-FEB-1991.

XX 02-AUG-1990; 90EP-0202113.

XX 04-AUG-1989; 89NL-0002010.

XX (NEZU-) NED INST ZUIVELONDE.

XX Vos PAU, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;

XX MPI; 1991-038622/06.

XX DR N-PSDB; AAQ10412.

XX Mutant protease gene(s) and protease(s) - derived from type I and

XX PT III protease genes from lactococcal strains, used in fermentation

XX PT foodstuffs and flavourings

XX PS Disclosure; Fig 1(1-7)+5(b); 29 pp; English.

XX The mutant protease having new cleavage specificities is

XX CC obtained by replacing three amino acids.

XX CC This mutant may then be used to prepare hybrid proteases,

XX CC the fusion being between a type I and a type III protease of

XX CC L.lactis Wg2 and SK11.

XX CC The product has modified properties, e.g. thermostability,

XX CC alkaline/acid pH stability, oxidative stability, autoprocolysis etc.,

XX CC compared to the parent protease(s). The proteases can be used for

XX CC preparing products (butter cheese, human and animal foodstuffs)

XX CC prepared with the aid of lactic acid bacteria.

XX CC See also AAQ10411-17 and AAQ10870-71.

XX CC

XX Sequence 1962 AA;

Query Match 21.6%; Score 951.5; DB 12; Length 1962;

Best Local Similarity 26.0%; Pred. No. Se-05; Indels 139; Gaps 33;

Matches 181; Conservative 133; Mismatches 243;

Qy 1 YPV--VLADTSSSEDALNISDEKEVAENKEKENHSHAMETSQDFEKKTAIVKEKEVVS 58

Db 24 LPVGEIQAKAISQQTIGSSLANVTAA-----TAKQATD-----TTA-----AT 64

Qy 59 KNPVIDNNTSNEAKIKKEBNSNKSQGDYTFVNNKNTENPKEDKVVVYIAEFKESGEK 118

Db 65 TNOAIAITOLAAKGIDYNNKLNKVKQODIYVDVIVQM--SAAPASENGILRTDVSSTAIEIOE 123

Qy 119 AIKELSLIKNTKVLV-----TYDRIFNAGSAIETTPDNLKIQIEGICISVERAOK 168
 Db 124 TNKVIYAAQASVKAIVEGVTQGTAGESYGVVNGFSTKVRVVDIPKIQIAGVKTIVLAKV 183
 Qy 169 VQPMNHARKKEIGVEA-IDYLSINAPFGNPFGRGVINISITGDTYRKAIRIDDA 227
 Db 184 YPTDAKANSMANVOAVWSNY-----KYVGEGVIVSVIDSGIDPTKDKRLSLDK 233
 Qy 228 KASRPFKEKL-KGTDK---NYWLSDKIPHAFNY--YNGCKITVEKYDDGRDYPFHGMHI 282
 Db 234 DV--KLTKSDVEKFTDVKHGRYNSKVPYGFNADNDITF---DDKVD--EONGHNV 285
 Qy 283 AGIAGNDTEODIXNFNGIDIAENAOIFSYKMSDAG--SGFAGDETFHAIEDSIKANV 341
 Db 286 AGIIGANGTGDPA--KSVGVAPAEQDLAMKVPNSDTPSGPSATVSAIEDSAKIGA 343
 Qy 342 DIVSVSSG-FTGTGLVEKTIQOAIRLRKAGIPNVVATGNYATSSSSWDLVANNHLKM 400
 Db 344 DVLNMSISGNSGNOTLEDPELAIVQANESGTAIVISAGNSGTGSATEG--VVKDYGL 401
 Qy 401 TDTGNVTRTAHEDAIVASAKNOTVERDKVNI-----GGSFKYRN---IGAFEDK 449
 Db 402 QDNEMVSGPGTSRGATTVASAEHTDVTQAVITDGTGLQIGPFTQLSSHDFTGS-FDQ 460
 Qy 450 SKITTNEGTAPAPSKLKFVYIGKQDODLIGDLRGKIAVMDRI-YTKDLNAPFKAMDK 508
 Db 461 KKFYIVVDASGN-----LSKGLADYTA-DAKGKIAIVKRGESFDDKQXQAQA-- 509
 Qy 509 GARAIMVYNTVYVNRDNTLPLPMGYEADGTYSVQFSISGDGVKL--NMNINPDCKT 566
 Db 510 GAAGLIIVNT-----DGTATPMT--SIALTTFPFGLSSVGOQLVDVWTAHPDQSL 560
 Qy 567 EVKRNKNEDEPKDKE-QYPIIDMESFNSNKNVDEKEDPKFAPDTDKELYKEDIYIPA 625
 Db 561 GVKIT---LAMPNQKYTEDEKYS-----DF----- 582
 Qy 626 GSTSGPRIDLLKPDVSAPGKNIKSTLVNIGKST 661
 Db 583 --TSYGPVSNLSFKPDITAPGGINIMSTON--NNGYTN 615

RESULT 12

AA010560
 ID AAR10560 standard; Protein; 1962 AA.

XX AAR10560;

XX 15-APR-1991 (first entry)

XX Mutant protease (K138N).

XX Mutant protease gene; fermentation; foodstuff; flavouring;

XX KM lactic acid bacteria.

XX OS Lactococcus lactis SK11.

XX Key Location/Qualifiers

FT Peptide 1..187

FT Protein /label= signal_peptide

FT Protein /label= mature_protein

XX EP411715-A.

XX 06-FEB-1991.

XX 02-AUG-1990; 90EP-0202113.

XX 04-AUG-1989; 89NL-0002010.

XX (NEZU-) NED INST ZUIVELONDE.

XX Vos PAU, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;


```

Db      124 TNKVIAAQAASVKAAVEQTQTAGESYGVVNGFSTKRVVDIPKLQIAGVKTVTLAKV 183
Qy      169 VQPMNNAHKEIGVEA-IDYLKSNAPFGNPFDRGVNISNIDGCTDYYRKAARIDDA 227
Db      184 YPPTDAKANSNANVOAVWSNY-----KYKGEQTVVSVISGIDPTEHKMRLSDK 233
Qy      228 KASRPFKEDL-KGTDK---NYWLSDKIPHAFNY-YNGSGKTVKEKYDDGRDYPDPHGMI 282
Db      234 DV--KLTSDEVEKFTDVKHGRYFNRSKYPYGFENADNDIT--DDKVD--EGHGHV 285
Qy      283 AGILAGNDEODIKNFNGIDGIAPNAQIFSYKMSDAG-SGFAGDETFHAIEDSIKINV 341
Db      286 AGIIGANGTGDPA--KSVGVAPPAQLAMKVFNSDPTSNANTGSATVVAIEDSAKIGA 343
Qy      342 DVSVSSGF-TGTGLVEKYMQAIRALKAGIPMVVATGNVATSSSSSWDLVANNHLM 400
Db      344 DVLNMSLGSNGQTLDEPDLAAVQNNESGTAIVASNGSGTSATEG--VKKDYGL 401
Qy      401 TDGTGNVTRTAHEDAIIVASAKNQTVEFDKNI-----GGSEFKYRN--IGAFEDK 449
Db      402 QDNEMVSGPGRSRAATTVASAENDVITQAVTITDGTGLQGPETIQLSSHDFTGS-FDQ 460
Qy      450 SKITTNEGTAKPSKLFVYIGKGDDDLGLDLRGKIAVNDRI-YTKDLKNAFKKAMD 508
Db      461 KKFYIVKADSGN-----LSKGLADYTA-DAKGIKAIYKGEFSDDKQKVAQA-- 509
Qy      509 GARAIMVNTVNYVNRDNTLPMAGYEADGTSQVFSISGDDGVKL--WNMINPDKT 566
Db      510 GAAGLIIVNT-----DGTATPMT--SIALTTFFPFGLSVVGQKLVDWVTAHPD 560
Qy      567 EVKRNKNEKDFDKLE-QYYPIDMESFNSNKNVGDGEKIDPKFAPDTKELYKEDI 625
Db      561 GVKIT--LAMPNQKYTEDKS-----DF----- 582
Qy      626 GSTSMGPRIDLLKPDVSAPGKNIKSTLVINGKST 661
Db      583 --TSYGPSNLSFKPDITAPGGINWSTQN--NNGYTN 615

RESULT 14
AAR10563
ID AAR10563 standard; Protein; 1962 AA.
XX
AC AAR10563;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (K748T).
XX
KM Mutant protease gene; fermentation; foodstuff; flavouring;
XX
KM Lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
XX
XX Key Location/Qualifiers
XX Peptide 1..187
XX FT /label= signal_peptide
XX FT Protein 188..1962
XX FT /label= mature_protein

EP411715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.
XX
XX Voe Pau, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;

```

```

DR      WPI; 1991-038622/06.
DR      N-PSDB; AAQ10417.
XX
XX      Mutant protease gene(s) and protease(s) - derived from type I and
PT      III protease genes from lactococcal strains, used in fermentation
PT      foodstuffs and flavourings
XX
XX      Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
XX      The mutant protease K748T having new cleavage specificities is
CC      obtained by carrying out single amino acid substitutions.
CC      This mutant may then be used to prepare hybrid proteases,
CC      the fusion being between a type I and a type III protease of
CC      L.lactis Wg2 and SK11.
CC      The product has modified properties, e.g. thermostability,
CC      alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC      compared to the parent protease(s). The proteases can be used for
CC      preparing products (butter cheese, human and animal foodstuffs)
CC      prepared with the aid of lactic acid bacteria.
CC      See also AAQ10411-17 and AAQ10870-71.
XX
XX      Sequence 1962 AA:
SQ
Query Match 21.6%; Score 948.5; DB 12; Length 1962;
Best Local Similarity 25.9%; Pred. No. 5.6e-05;
Matches 180; Conservative 134; Mismatches 243; Indels 139; Gaps 33;

Qy      1 YPV-VIADTSSSDALNISPEKVAENKEHNIHSMETSOPFEKKTAVIYEKYS 58
Db      24 LPVEIGAKKAISQOTLGSSLANVTAA-----TAKQAAVT-----AT 64
Qy      59 KNPYIDNNTSNEAKIKIEENSKSQDYDTSFVNKNTPENPKEDKVVYIAFKDSEGEK 118
Db      65 TNQAIATQLAKKGIIDYKLNKVVQODIYVDIVM-SAAPSENGILRTDYSSTAEIQE 123
Qy      119 AIKELSLKTKTVLY-----TYDRIFNGSAIETTPNDLKIQIEGSISSVERAQK 168
Db      124 TNKVIAAQAASVKAAVEQTQTAGESYGVVNGFSTKRVVDIPKLQIAGVKTVTLAKV 183
Qy      169 VQPMNNAHKEIGVEA-IDYLKSNAPFGNPFDRGVNISNIDGCTDYYRKAARIDDA 227
Db      184 YPPTDAKANSNANVOAVWSNY-----KYKGEQTVVSVISGIDPTEHKMRLSDK 233
Qy      228 KASRPFKEDL-KGTDK---NYWLSDKIPHAFNY-YNGSGKTVKEKYDDGRDYPDPHGMI 282
Db      234 DV--KLTSDEVEKFTDVKHGRYFNRSKYPYGFENADNDIT--DDKVD--EGHGHV 285
Qy      283 AGILAGNDEODIKNFNGIDGIAPNAQIFSYKMSDAG-SGFAGDETFHAIEDSIKINV 341
Db      286 AGIIGANGTGDPA--KSVGVAPPAQLAMKVFNSDPTSNANTGSATVVAIEDSAKIGA 343
Qy      342 DVSVSSGF-TGTGLVEKYMQAIRALKAGIPMVVATGNVATSSSSSWDLVANNHLM 400
Db      344 DVLNMSLGSNGQTLDEPDLAAVQNNESGTAIVASNGSGTSATEG--VKKDYGL 401
Qy      401 TDGTGNVTRTAHEDAIIVASAKNQTVEFDKNI-----GGSEFKYRN--IGAFEDK 449
Db      402 QDNEMVSGPGRSRAATTVASAENDVITQAVTITDGTGLQGPETIQLSSHDFTGS-FDQ 460
Qy      450 SKITTNEGTAKPSKLFVYIGKGDDDLGLDLRGKIAVNDRI-YTKDLKNAFKKAMD 508
Db      461 KKFYIVKADSGN-----LSKGLADYTA-DAKGIKAIYKGEFSDDKQKVAQA-- 509
Qy      509 GARAIMVNTVNYVNRDNTLPMAGYEADGTSQVFSISGDDGVKL--WNMINPDKT 566
Db      510 GAAGLIIVNT-----DGTATPMT--SIALTTFFPFGLSVVGQKLVDWVTAHPD 560
Qy      567 EVKRNKNEKDFDKLE-QYYPIDMESFNSNKNVGDGEKIDPKFAPDTKELYKEDI 625
Db      561 GVKIT--LAMPNQKYTEDKS-----DF----- 582
Qy      626 GSTSMGPRIDLLKPDVSAPGKNIKSTLVINGKST 661

```

```

Db      583 --TSYGPVSNLSFKPDITAPGGNIWSTQN-NNGYTN 615
RESULT 15
AAAR10562
ID      AAR10562 standard; Protein; 1959 AA.
XX
XX
XX      AAR10562;
XX
XX      15-APR-1991 (first entry)
XX
XX      Mutant protease (delta137-139).
DE
XX
XX      Mutant protease gene; fermentation; foodstuff; flavouring;
KW      lactic acid bacteria.
XX
XX      Lactococcus lactis SK11.
XX
XX      Key      Location/Qualifiers
FT      Peptide      1..187
FT      Protein      /label= signal_peptide
FT      FT          188..1959
FT      FT          /label= mature_peptide
XX
XX      EP411715-A.
XX
XX      06-FEB-1991.
XX
XX      02-AUG-1990; 90EP-0202113.
XX
XX      04-AUG-1989; 89NL-0002010.
XX
XX      (NEZU-) NED INST ZUIVELONDE.
XX
XX      Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX      WPI; 1991-038622/06.
XX      N-PSDB; AAQ104116.
XX
XX      Mutant protease gene(s) and protease(s) - derived from type I and
XX      III protease genes from lactococcal strains, used in fermentation
XX      foodstuffs and flavourings
XX
XX      Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
XX      The mutant protease having new cleavage specificities is
XX      obtained by deleting three amino acids.
XX      This mutant may then be used to prepare hybrid proteases,
XX      the fusion being between a type I and a type III protease of
XX      L.lactis Wg2 and SK11.
XX      The product has modified properties, e.g. thermostability,
XX      alkaline/acid pH stability, oxidative stability, autolysis etc.,
XX      compared to the parent protease(s). The proteases can be used for
XX      preparing products (butter cheese, human and animal foodstuffs)
XX      prepared with the aid of lactic acid bacteria.
XX      See also AAQ10411-17 and AAQ10870-71.
XX
XX      Sequence 1959 AA;

```

```

Query Match      21.5%; Score 947; DB 12; Length 1959;
Best Local Similarity 25.8%; Pred. No. 5.9e-05;
Matches 179; Conservative 135; Mismatches 241; Indels 140; Gaps 33;

QY      1 YPV--VLADTSSSEDALNISDEKVAENKEKHENIHSAMETSQDFKEKTAIVIKEWS 58
DB      24 LPVGEIQAKAATISQQTGLSSSLANTVTAA-----TAKQAATD---TTA-----AT 64
QY      59 KNPVINDNTSNEBAKIKENSNSQGDYDTSFVNKNTENPKKEDKVVVIAEFKDKESKEK 118
DB      65 TNQAIATQLAAKGIIDYNKLNKVVQOQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQOE 123
QY      119 AIKELSSLNKTPLY-----TYDRIFNGSAIETTPDNLDKIKIEGISSVERAQK 168

```

Search completed: October 9, 2003, 15:51:10
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2003, 15:48:38 ; Search time 85 Seconds
(without alignments)
1239.935 Million cell updates/sec

Title: US-09-590-991-6

Perfect score: 4398

Sequence: 1 YPVVADTSSSEDALNISDK.....FGKNIKSTLVINGKSTGY 664

Scoring table:

BIOSUN30
Gapop 8.0 , Gapext 0.1

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq_19Jun03.*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4398	100.0	664	22 AAB48342	S. pneumoniae Sp12
2	4398	100.0	2120	21 AAY81710	Streptococcus pneu
3	4381.7	99.6	2140	24 ABU01020	S. pneumoniae type
4	4171	94.8	1007	19 AAW61246	Streptococcus pneu
5	4171	94.8	1007	23 ABP54664	S. pneumoniae Sp12
6	1189.3	27.0	2366	19 AAW68388	Clostridium diffic
7	1188.3	27.0	2366	17 AAR95011	C. difficile toxin
8	1186.4	27.0	3132	22 AAB51367	Japanese Yam mosai
9	1175.3	26.7	6619	22 ABB23329	Novel human diagno

10	1173	26.7	5909	22 ABB23295	Novel human diagno
11	1167.8	26.6	7107	22 ABB58144	Drosophila melanog
12	1165.4	26.5	2835	23 ABB98574	Dextran saccharase
13	1165.2	26.5	31267	24 ABB74786	Human RGS11 protei
14	1164.4	26.5	4345	22 ABB66417	Drosophila melanog
15	1161.5	26.4	26926	22 AAU05396	Human titlin (cone
16	1156.9	26.3	2946	22 ABB25790	Novel human diagno
17	1156.1	26.3	5464	22 ABB11810	Novel human diagno
18	1153.3	26.2	3781	22 ABB73291	Novel human diagno
19	1151.9	26.2	4643	22 ABB71609	Drosophila melanog
20	1151.8	26.2	2710	17 AAR95016	C. difficile toxin
21	1151.8	26.2	2710	19 AAW68387	Clostridium diffic
22	1151.4	26.2	3909	23 ABB3661	mRNA for hPac prot
23	1151.4	26.2	4590	23 ABB68608	Human pancreatic c
24	1151.4	26.2	4590	24 AAO26790	Human cadherin (CA
25	1151.4	26.2	4590	24 AAO26791	Human cadherin (CA
26	1151.4	26.2	4591	22 ABB22977	Novel human diagno
27	1150.9	26.2	2233	24 ABU01029	S. pneumoniae type
28	1150.9	26.2	2234	21 AAY61502	Streptococcus pneu
29	1148.8	26.1	3269	24 ABB54436	Papaya leaf-dictor
30	1147.6	26.1	15281	15 AAR44928	T. niyeum Cyclospo
31	1147.4	26.1	4081	22 ABB62958	Drosophila melanog
32	1147.3	26.1	3457	15 AAR62504	Large polypeptid
33	1142.5	26.0	3443	20 AAW84559	Polypeptid encode
34	1142.1	26.0	2681	24 ABB19025	Pathogen specific
35	1140.4	25.9	4796	22 ABB58665	Drosophila melanog
36	1138.9	25.9	4555	23 AAM52106	Rat fat 3 protein
37	1134.7	25.8	4599	24 ABB56837	Human LRP1B protei
38	1133.5	25.8	4472	22 ABB60101	Drosophila melanog
39	1133.3	25.7	3457	20 AAW84560	Polypeptid encode
40	1131.7	25.7	4636	22 AAE11937	Human CG168 (or CS
41	1130.2	25.7	3542	22 ABB62142	P. falciparum PCR3
42	1130.2	25.7	7718	22 ABB11811	Novel human diagno
43	1128.8	25.7	3097	22 ABB2967	Drosophila melanog
44	1127.1	25.6	4152	22 AAG64992	Shrimp white spot
45	1126.6	25.6	5635	23 ABB60991	Novel human protei

ALIGNMENTS

RESULT 1	AAB48342	standard; Protein: 664 AA.
AC	AAB48342;	
DT	20-APR-2001 (first entry)	
XX		
XX		S. pneumoniae Sp12 polypeptide.
XX		Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW		bronchial; lung; blood; infection; immune response; immunotherapy;
KW		antibacterial; auditory; vaccine.
XX		
OS		Streptococcus pneumoniae.
PN	WO200076540-A2.	
PD	21-DEC-2000.	
PF	09-JUN-2000; 2000MO-US15925.	
PR	10-JUN-1999; 99US-0138453.	
PA	(MED1-) MED IMMUNE INC.	
PI	Adamou JE, Choi GH;	
XX	WPI; 2001-112197/12.	
DR	N-PSDB; AAC84741.	
XX		
PT		New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
XX blood infections
XX
PS Claim 4; Page 47-50; 54pp; English.
XX
CC The invention relates to novel immunogenic polypeptides, Spi28 and Spi30
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC the treatment and prevention of pneumococcal infections, particularly
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC bronchial, lung or blood infections. The antigens are used as immunogenic
CC agents to stimulate an immune response. The antisera and antibodies may
CC also be used in diagnosing and treating pneumococcal infections.
CC Recombinant polypeptides serve as a mechanism for stimulating production
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC as reagents in other processes such as affinity chromatography. The
CC present sequence represents the S. pneumoniae Spi28 polypeptide.
XX
SQ Sequence 664 AA;

Query Match 100.0%; Score 4398; DB 22; Length 664;
Best Local Similarity 100.0%; Pred. No. 3.7e-73;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YPVVLADTSSSEDALNISDKVAENKEKHENIHSAMETSDQFKEKKTAVIKEKEVWSKN 60
DB 1 YPVVLADTSSSEDALNISDKVAENKEKHENIHSAMETSDQFKEKKTAVIKEKEVWSKN 60
QY 61 PVIDNNTSNEEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120
DB 61 PVIDNNTSNEEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120
QY 121 KELSLKNTKVLTYDRIFFNGSAIETTPDNLDKIKQIEGSISSVERAQKQVPMNHARKEI 180
DB 121 KELSLKNTKVLTYDRIFFNGSAIETTPDNLDKIKQIEGSISSVERAQKQVPMNHARKEI 180
QY 181 GVEEAIDVLKSNAPGKNGFGRGVISNIDGTDRYRKAMRIDDAKAMRFKEDLKG 240
DB 181 GVEEAIDVLKSNAPGKNGFGRGVISNIDGTDRYRKAMRIDDAKAMRFKEDLKG 240
QY 241 TDKNYWLSDKIPAFNYNGGKITVEKYDDGRDYDPHGMHAGIAGNDETDQDKNFNG 300
DB 241 TDKNYWLSDKIPAFNYNGGKITVEKYDDGRDYDPHGMHAGIAGNDETDQDKNFNG 300
QY 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETFWPHAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
DB 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETFWPHAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
QY 361 WQAIRALKAGIPMVVATGNYATSASSSSDWLVANNHLKMTDTGNVTRTAHEDAIAVAS 420
DB 361 WQAIRALKAGIPMVVATGNYATSASSSSDWLVANNHLKMTDTGNVTRTAHEDAIAVAS 420
QY 421 AKNQTVFEDKVNIGGESFYRNIGAFDDKSKITTNEDGTGKAPSKLKFVYIGKQDQDLIG 480
DB 421 AKNQTVFEDKVNIGGESFYRNIGAFDDKSKITTNEDGTGKAPSKLKFVYIGKQDQDLIG 480
QY 481 LDLRGKIAVMRIYTKDLKNAPFKAMDKGARAIMVNTVNYNRDNTWELPAMGYEADBG 540
DB 481 LDLRGKIAVMRIYTKDLKNAPFKAMDKGARAIMVNTVNYNRDNTWELPAMGYEADBG 540
QY 541 TKSQVFSISGDDGVKLWNNINPDKTEVKRNKEDFKDKLEQYYPIDMESFNNSKPNVGD 600
DB 541 TKSQVFSISGDDGVKLWNNINPDKTEVKRNKEDFKDKLEQYYPIDMESFNNSKPNVGD 600
QY 601 EKEIDFKFAPDTDKELYEDIIIVPAGSTSWGPRIDLKLPDVSAPFGKNIKSTLNIVNGKS 660
DB 601 EKEIDFKFAPDTDKELYEDIIIVPAGSTSWGPRIDLKLPDVSAPFGKNIKSTLNIVNGKS 660
QY 661 TYGY 664
DB 661 TYGY 664

RESULT 2

AAV81710
ID AAV81710 standard; Protein; 2120 AA.
XX
AC AAV81710;
XX
DT 02-JUN-2000 (first entry)
XX
DE Streptococcus pneumoniae protein sequence ID3.
XX
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
OS Streptococcus pneumoniae.
XX
PN WO200006738-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB02452.
XX
PR 27-JUL-1999; 98GB-0016336.
XX
PR 19-MAR-1999; 99US-0125329.
XX
PA (MICR-) MICROBIAL TECHNIQS LTD.
XX
PI Le Page RWF, Wells JM, Hanniffy SB, Hanabro PM;
XX
XX WPI; 2000-195301/17.
DR N-PSDB; AAZ91806.
XX
XX Streptococcal proteins and polynucleotides useful for diagnosis,
PT treatment and prophylaxis of bacterial infections
XX
XX Claim 2; Page 41-42; 76pp; English.
XX
CC This sequence represents a Streptococcus pneumoniae protein of the
CC invention. The proteins (or their homologues, derivatives and/or
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
CC compositions comprising the proteins are useful as vaccines and also in
CC diagnostic assays. The sequences are useful for the detection or
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
CC with them. Agents capable of antagonising, inhibiting or interfering with
CC the function or expression of the protein or polypeptide are useful in
CC medical compositions in the treatment or prophylaxis of S. pneumoniae
CC infection. As the sequences can be used to treat S. pneumoniae infection,
CC they can be used to treat bacterial pneumonia, which has high rates in
CC young children, the elderly, and in patients with predisposing conditions
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC or with immunosuppressive disorders, especially AIDS. They can also be
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC meningitis.
XX
SQ Sequence 2120 AA;
Query Match 100.0%; Score 4398; DB 21; Length 2120;
Best Local Similarity 100.0%; Pred. No. 3.3e-71;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YPVVLADTSSSEDALNISDKVAENKEKHENIHSAMETSDQFKEKKTAVIKEKEVWSKN 60
DB 1 YPVVLADTSSSEDALNISDKVAENKEKHENIHSAMETSDQFKEKKTAVIKEKEVWSKN 60
QY 61 PVIDNNTSNEEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120
DB 61 PVIDNNTSNEEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120
QY 121 KELSLKNTKVLTYDRIFFNGSAIETTPDNLDKIKQIEGSISSVERAQKQVPMNHARKEI 180
DB 121 KELSLKNTKVLTYDRIFFNGSAIETTPDNLDKIKQIEGSISSVERAQKQVPMNHARKEI 180

QY 181 GVEEAIDVLSKINAPFGKDFGRGMVINSIDTGTDRHKAARIDDDAKASNRFKKEDLKG 240
DB 181 GVEEAIDVLSKINAPFGKDFGRGMVINSIDTGTDRHKAARIDDDAKASNRFKKEDLKG 240
QY 241 TDKNWLSDKIPHANVYNGGKITVEKYDDGRDYPDPHGMHAGIAGLNDTEODIKNNG 300
DB 241 TDKNWLSDKIPHANVYNGGKITVEKYDDGRDYPDPHGMHAGIAGLNDTEODIKNNG 300
QY 301 IDGIAFNQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
DB 301 IDGIAFNQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
QY 361 WQAIKALRKAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDGTGVTFTAHEDAIAVAS 420
DB 361 WQAIKALRKAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDGTGVTFTAHEDAIAVAS 420
QY 421 AKNQVFEFDKVNIGESFKNYKIGAFDPKSKITTNEDEGTAKPSKLFYIGKQDQDLIG 480
DB 421 AKNQVFEFDKVNIGESFKNYKIGAFDPKSKITTNEDEGTAKPSKLFYIGKQDQDLIG 480
QY 481 LDLRGKIVMDRIYTKDLKNAFKKAMDGARAIMVNVNRYNRMNTELPAMGYEADG 540
DB 481 LDLRGKIVMDRIYTKDLKNAFKKAMDGARAIMVNVNRYNRMNTELPAMGYEADG 540
QY 541 TKSQVFSISGDDGVKLMNMINDEKTEVYRNKKEPDFKLEQYYPIDMESEFNSKNPNVGD 600
DB 541 TKSQVFSISGDDGVKLMNMINDEKTEVYRNKKEPDFKLEQYYPIDMESEFNSKNPNVGD 600
QY 601 EKEIDFKAPPTDKELYKEDIIVPAGSTSWGPRIDLKLPDVSAPGKNIKSTLVYNGKS 660
DB 601 EKEIDFKAPPTDKELYKEDIIVPAGSTSWGPRIDLKLPDVSAPGKNIKSTLVYNGKS 660
QY 661 TYGY 664
DB 661 TYGY 664

RESULT 3
ABU01020
ID ABU01020 standard; Protein; 2140 AA.

XX AC ABU01020;
XX DT 11-FEB-2003 (first entry)
XX DE S. pneumoniae type 4 strain protein from coding region #590.
XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;
XX KW ear infection; antiinflammatory; antibacterial; immunostimulant;
XX KW auditory; respiratory; gene therapy; vaccine.
XX OS Streptococcus pneumoniae type 4 strain.
XX PN WO200277021-A2.
XX PD 03-OCT-2002.
XX PF 27-MAR-2002; 2002WO-1B02163.
XX PR 27-MAR-2001; 2001GB-0007658.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Masignani V, Tettelin H, Frazer C;
XX DR MPI; 2003-040579/03.
XX DR N-PSDB; ABX06302.
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection

XX Claim 1; SEQ ID No 1180; 56bp; English.
XX
PS The invention relates to a protein comprising or having at least 50%
SS identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB856454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2469 proteins expressed by the identified coding regions from the
CC genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2140 AA;

Query Match 99.6%; Score 4381.7; DB 24; Length 2140;
Best Local Similarity 99.6%; Pred. No. 76-71; 0; Indels 3; Gaps 2;
Matches 664; Conservative 0; Mismatches 0;

QY 1 Y-P--VVLADTSSSEDLNISDEKEVAENKEHENHISAMETSODPFKKTAIVKEKEVV 57
DB 19 YAPNEVVLADTSSSEDLNISDEKEVAENKEHENHISAMETSODPFKKTAIVKEKEVV 78
QY 58 SKNPVIDNNTSNEBAKIKENSNSKSGDYTDSFVNKNTENPKKDKVYVIAEPKDSGE 117
DB 79 SKNPVIDNNTSNEBAKIKENSNSKSGDYTDSFVNKNTENPKKDKVYVIAEPKDSGE 138
QY 118 KAIKELSLKNTKVLVYDRIFNGSAIETTPDNLDKI KOLEGISVERAOKVQPMNHR 177
DB 139 KAIKELSLKNTKVLVYDRIFNGSAIETTPDNLDKI KOLEGISVERAOKVQPMNHR 198
QY 178 KEIGVEEAIDVLSKINAPFGKDFGRGMVINSIDTGTDRHKAARIDDDAKASNRFKKED 237
DB 199 KEIGVEEAIDVLSKINAPFGKDFGRGMVINSIDTGTDRHKAARIDDDAKASNRFKKED 258
QY 238 LKGTDKNWLSDKIPHANVYNGGKITVEKYDDGRDYPDPHGMHAGIAGLNDTEODIKN 297
DB 259 LKGTDKNWLSDKIPHANVYNGGKITVEKYDDGRDYPDPHGMHAGIAGLNDTEODIKN 318
QY 298 FNGIDGIAFNQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLV 357
DB 319 FNGIDGIAFNQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLV 378
QY 358 EKYWQAIKALRKAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDGTGVTFTAHEDAIA 417
DB 379 EKYWQAIKALRKAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDGTGVTFTAHEDAIA 438
QY 418 VASAKNQVFEFDKVNIGESFKNYKIGAFDPKSKITTNEDEGTAKPSKLFYIGKQDQD 477
DB 439 VASAKNQVFEFDKVNIGESFKNYKIGAFDPKSKITTNEDEGTAKPSKLFYIGKQDQD 498

478	Qy	LIGLDLRGKIAVMDRIYTKDLKNAPKKAMDGSAIMVNTVNYNRDNWTELPAMGYEA	537
499	Db	LIGLDLRGKIAVMDRIYTKDLKNAPKKAMDGSAIMVNTVNYNRDNWTELPAMGYEA	558
538	Qy	DGTSKQVFSISGDDGVKLWNMINPDKKTVEYKRNKEDFKDKLEQYYPIOMESFNSNKP	597
559	Db	DGTSKQVFSISGDDGVKLWNMINPDKKTVEYKRNKEDFKDKLEQYYPIOMESFNSNKP	618
598	Qy	VGDEKEIDPFKFPADTDKELYKEDIIVPAGSTSWGPRIDLKLLKPDVSPAGPKNIKSTLNVIN	657
619	Db	VGDEKEIDPFKFPADTDKELYKEDIIVPAGSTSWGPRIDLKLLKPDVSPAGPKNIKSTLNVIN	678
658	Qy	GKSTGYG 664	
679	Db	GKSTGYG 685	

RESULT 4
AAW61246
ID AAW61246 standard; Protein; 1007 AA.
XX
XX AAW61246;
XX AC
XX AC
DT 02-OCT-1998 (first entry)

AA Nucleic acid encoding antigenic peptide(s) from *Streptococcus*
PT
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis

PS Claim 11; Page 92-93; 118pp; English.

The present sequence represents a protein from *Streptococcus pneumoniae*. The nucleic acid sequence encoding the *Streptococcus pneumoniae* protein can be useful in vaccines for inducing protective antibodies against *Streptococcus pneumoniae*, for treatment or prevention of infection e.g. pneumonia, otitis media or meningitis. Probes based on the nucleic acid sequence are used to detect *Streptococcus* infection (by usual hybridisation or amplification methods), also for isolating *Streptococcus* genes or their allelic variants. The protein can be used similarly to detect specific antibodies in standard immunoassays, especially for diagnosing or monitoring infections. Antibodies which bind the protein are used to detect corresponding antigens, to purify the protein and for passive immunisation (optionally coupled to a toxin). Vaccines are administered e.g. by injection, orally or through the skin, typically at 0.01-1000 (especially 10-300) $\mu\text{g/ml}$ per dose.

Sequence	1007 AA;
SQ	

Query Match	94.8%	Score 4171	DB 19	Length 1007
Best Local Similarity	100.0%	Pred. No. 4.4e-68		
Matches 627	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy	38	ETSDQPEKKTAVI	KEKEVVSKNPV	DNNTSNEAKI	KEENS	NKSQGDY	TDSFVN	KNTEN	97		
Db	1	ETSDQPEKKTAVI	KEKEVVSKNPV	DNNTSNEAKI	KEENS	NKSQGDY	TDSFVN	KNTEN	60		
Qy	98	PKKEDKVVI	TAEPKDKESGEKA	IKELSS	LKNTKVLY	YDRI	FNGSA	LETTPDNLDKIKOI	157		
Db	61	PKKEDKVVI	TAEPKDKESGEKA	IKELSS	LKNTKVLY	YDRI	FNGSA	LETTPDNLDKIKOI	120		
Qy	158	EGISSVERAQV	OPMNHARKEI	GVBEA	IDYLS	INAPFG	KNPFGR	GMVINSIDTGTDYR	217		
Db	121	EGISSVERAQV	OPMNHARKEI	GVBEA	IDYLS	INAPFG	KNPFGR	GMVINSIDTGTDYR	180		
Qy	218	HKAWR	TDODAKASMR	FKKEDL	KGTCKNY	WLSDKI	PHAFNY	YNGKITEVKYDDGRDYDP	277		
Db	181	HKAWR	TDODAKASMR	FKKEDL	KGTCKNY	WLSDKI	PHAFNY	YNGKITEVKYDDGRDYDP	240		
Qy	278	HGMH	IAGILAGN	TEQDI	KNFN	FGIDGI	APNAQIF	SYKMYSDAGSGFAGDETMFHAIEDSI	337		
Db	241	HGMH	IAGILAGN	TEQDI	KNFN	FGIDGI	APNAQIF	SYKMYSDAGSGFAGDETMFHAIEDSI	300		
Qy	338	KHNVDVV	VSSGGTGTGL	VGEK	VQWAI	RALR	KAGIPMV	VATGNVAT	SASSSSWDLVANNH	397	
Db	301	KHNVDVV	VSSGGTGTGL	VGEK	VQWAI	RALR	KAGIPMV	VATGNVAT	SASSSSWDLVANNH	360	
Qy	398	LKMTD	TGNVTR	TAAHEDA	IAVAS	AKNOTV	BEFDK	VNIGGESFKYRNIGAF	FPDKSKITTNED	457	
Db	361	LKMTD	TGNVTR	TAAHEDA	IAVAS	AKNOTV	BEFDK	VNIGGESFKYRNIGAF	FPDKSKITTNED	420	
Qy	458	GTKAP	SKLFVYIG	KGQDD	LIGL	DRGKIA	VMDRIY	TKOLKN	AFKKAMDKGARAIMVYN	517	
Db	421	GTKAP	SKLFVYIG	KGQDD	LIGL	DRGKIA	VMDRIY	TKOLKN	AFKKAMDKGARAIMVYN	480	
Qy	518	TVNY	YARDN	WTELP	PAMGYE	ADG	TKSOV	FISIGDDG	VKLMNINPDKKTEVKRNKKE	577	
Db	481	TVNY	YARDN	WTELP	PAMGYE	ADG	TKSOV	FISIGDDG	VKLMNINPDKKTEVKRNKKE	540	
Qy	578	DKLEQ	YYPID	MESFN	SKPNV	GEDE	IDFK	FAPDTDKEL	YKEDIIVPAG	STSWGPRIDLL	637
Db	541	DKLEQ	YYPID	MESFN	SKPNV	GEDE	IDFK	FAPDTDKEL	YKEDIIVPAG	STSWGPRIDLL	600
Qy	638	LKPD	V	SAPG	KNIK	STL	N	V	NGKSTYGY	664	
Db	601	LKPD	V	SAPG	KNIK	STL	N	V	NGKSTYGY	627	

RESULT 5
ABP54664

ABP54664
ID ABP54664 standard; Protein; 1007 AA.

AC ABP54664;

XX
DT 04-SEP-2002 (first entry)XX
DE
DE
S. pneumoniae SP122 protein sequence SEQ ID NO:216.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

XX
PN US2002061545-A1.

XX PD 23-MAY-2002.

XX PF 22-JAN-2001: 2001US-0765272.

XX
PR 30-OCT-1997: 97US-0961083.XX
PA (CHOI/) CHOI G H.

PA (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.

PA (BARA//) BARASH S C.

PA (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX
 XX
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,
 PI Rosen CA;
 DR WPI, 2002-479261/51.
 DR N-PSDB; ABQ84899.
 XX
 XX
 PT New Streptococcus pneumoniae antigens, useful for detecting
 PT Streptococcus and for preventing or attenuating disease caused by
 PT Streptococcus infection -
 PS
 PS Claim 11, Page 50; 70pp; English.
 XX
 XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
 CC which are used in an example from the present invention.
 CC
 SQ Sequence 1007 AA;
 Query Match 94.8%; Score 4171; DB 23; Length 1007;
 Best Local Similarity 100.0%; Pred. No. 4,4e-68;
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 38 ETSODPFEKKTAVIEKEVEVSNPVI DNNTSNEBAKIKENSNKSGOYDTSFVNKNEN 97
 Db 1 ETSQDFEKTAVIEKEVEVSNPVI DNNTSNEBAKIKENSNKSGOYDTSFVNKNEN 60
 Oy 98 PKKEDKVVYIAEFKKESEGEKAIKELSLKNTKVLTYTDRIFNGSAIETTPNDLKIKOI 157
 Db 61 PKKEDKVVYIAEFKKESEGEKAIKELSLKNTKVLTYTDRIFNGSAIETTPNDLKIKOI 120
 Oy 158 EGISSEVERAQQVQPMNTHARKEIGVEAIDYKSIINAPFGKNFDRGMVINSIDTGYR 217
 Db 121 EGISSEVERAQQVQPMNTHARKEIGVEAIDYKSIINAPFGKNFDRGMVINSIDTGYR 180
 Oy 218 HKAMIIDDDAKASMRFKKEDLKTGDKNWLSDKI PHAFYVYNGKITVEKYDGDGDP 277
 Db 181 HKAMIIDDDAKASMRFKKEDLKTGDKNWLSDKI PHAFYVYNGKITVEKYDGDGDP 240
 Oy 278 HGMHAGILAGNDTODIKNFNGIDGIA PNAQIFSYKNYSDAGSGFAGDETFFHAIEDSI 337
 Db 241 HGMHAGILAGNDTODIKNFNGIDGIA PNAQIFSYKNYSDAGSGFAGDETFFHAIEDSI 300
 Oy 338 KHNVDVSVSSGFTGTGLVEKEYWQAIRLRKAGI PMVATGNVATSASSSSMDLVANNH 397
 Db 301 KHNVDVSVSSGFTGTGLVEKEYWQAIRLRKAGI PMVATGNVATSASSSSMDLVANNH 360
 Oy 398 LKMDTGTGVTTRAHAEHDAIAVASAKQVTEPDKNVIIGESFYNIGAFPDKSKITTTED 457
 Db 361 LKMDTGTGVTTRAHAEHDAIAVASAKQVTEPDKNVIIGESFYNIGAFPDKSKITTTED 420
 Oy 458 GTKAPSKLKFVYIGKQODODLIGDLRGKIAMDRITTKDLKNAFKKAMDKARAIMVYN 517
 Db 421 GTKAPSKLKFVYIGKQODODLIGDLRGKIAMDRITTKDLKNAFKKAMDKARAIMVYN 480
 Oy 518 TVNYYNRDNWTELPMAGYEADGETSKSVFSISGDDGVKLMNINDDKTEVRKNKEDPK 577
 Db 481 TVNYYNRDNWTELPMAGYEADGETSKSVFSISGDDGVKLMNINDDKTEVRKNKEDPK 540
 Oy 578 DKLGEYVPIIDMESFNSKNPNVNGDEKEIDFKFAPDIDKLYKEDIIVPAGSTSMGRIDL 637
 Db 541 DKLGEYVPIIDMESFNSKNPNVNGDEKEIDFKFAPDIDKLYKEDIIVPAGSTSMGRIDL 600
 Oy 638 LKPDVSABQKNIKSTLVINGKSTGYG 664

Db 601 LKPDVSABQKNIKSTLVINGKSTGYG 627
 |||||
 RESULT 6
 AAW68388
 ID AAW68388 standard; Protein; 2366 AA.
 XX
 XX
 AC AAW68388;
 XX
 DT 07-DEC-1998 (first entry)
 XX
 DE Clostridium difficile toxin B.
 XX
 XX Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;
 KM pseudomembranous enterocolitis.
 XX
 XX Clostridium difficile.
 XX
 OS MO9808540-A1.
 XX
 PN 05-MAR-1998.
 PD
 XX
 PF 28-AUG-1997; 97WO-US15394.
 XX
 PR 28-AUG-1996; 96US-0704159.
 XX
 PA (OPHI-) OPHIDIAN PHARM INC.
 PI Thallely BS, Williams JA;
 XX
 DR WPI: 1998-230234/20.
 XX
 DR N-PSDB; AAV30561.
 XX
 PT Host cell containing recombinant expression vector encoding
 PT Clostridium botulinum type B or E toxin - useful to treat humans
 PT and other animals at risk of intoxication with clostridial toxin
 XX
 PS Example 18; Page 241-249; 428pp; English.
 XX
 CC This is the amino acid sequence of Clostridium difficile toxin B,
 CC deduced from the coding region (see AAV30561) of the toxin B gene.
 CC Fragments of the toxin B gene have been cloned into various
 CC prokaryotic expression systems, and assessed for the ability to
 CC express recombinant toxin B protein in E. coli. It would be
 CC advantageous to use simple and inexpensive prokaryotic expression
 CC systems to produce and purify high levels of recombinant toxin B
 CC for immunisation purposes. The invention specifically relates to
 CC recombinant proteins derived from Clostridium botulinum toxins
 CC (see AAW68389-400) and their use as immunogens for the production of
 CC vaccines and antitoxins.
 CC
 SQ Sequence 2366 AA;
 Query Match 27.0%; Score 1189.3; DB 19; Length 2366;
 Best Local Similarity 12.4%; Pred. No. 3.8e-09;
 Matches 292; Conservative 143; Mismatches 198; Indels 1717; Gaps 119;
 Oy 1 YP-----VV-----LADTSSSEDALNISDEKVAENK----- 27
 Db 34 YHNNSENVTEVEKYKLKDINSLTDIY-IDTYKSGGRKALKKFKFYLVTVEVLKNNNLT 92
 Oy 28 --EKHENTH----- 36
 Db 93 PVKE--NLHFWIGQINDTAINYINQKDVNSDVNVVFPYDSNAFLINTLTKTVESAI 150
 Oy 37 METSODPFE-----KKTAVI--KEKEV-----SKNP--VIDNT----- 67
 Db 151 NDTLESFRENLDNRFDYNNKFFRKQMOIYDKQNFINYAKORENPELIIIDIVKTYL 210
 Oy 68 SNEBAK-IKENSS-----NK-SQ----- 83
 Db 211 SNEYSKIDEIDELNTYIESLNKNTITQNSGNDVNRNFEFPKNGSSFNLVBEQELVERMNLAAASD 270

QY 84 -----GDYTD----- 88
 Db 271 ILRISALKEIGMYLDVDMPLGOIPDLFESIEKPSVTVDFWMTKLEAIMKYEYIPEY 330
 QY 89 -----SF-----VN----- 92
 Db 331 TSEHDMLEEVQSSFSVLASKSDKSEIFSSSLGDMASPLEVKIAPNSKGIINOGLISV 390
 QY 93 -----KXNENPK-KED----- 102
 Db 391 KDSYCSNLIVKQIENRYKILNLSNPAISEDNDFTNTTTFIDSIAMAEANADNGRFMMEL 450
 QY 103 -KVVVIAEPKKE-----SGEKA-----IKELS-----SLKNTKV----- 131
 Db 451 KYLRVGFPPDVKTINLSGPEAYAAAYQDILLMFKEGSMNHLIBADLRNFEISKNTISQ 510
 QY 132 -----LYTVD-----RI-PNGSAIETTPDNLDKIKOI-----EGISS 162
 Db 511 STEQEMASLWFFDDARAKAQPEEYKRNYPFEGSLGE--DDNLDPSQNIIVVDKEYLLEKISS 568
 QY 163 VERA----- 166
 Db 569 LARSSERGIHIVQLQDKISYEAACNLFAKTPYDSVLFQKNIEDSEIAYYINPGDGEI 628
 QY 167 QKV-----OPM-----MNHARKE-----IG-----VEEAIDYLK----- 190
 Db 629 QEIDKYKIPSIIDRPKIKLTFIGHGKDFNTDIFAGFDVDSLSTEIEAAIDLAKEDISP 688
 QY 191 -----SINA-----PFGK-----N 199
 Db 689 KSIEINLLGCNMFYSINVEETYP-GKLLLVKDKISELMPISQDSIIVSANQYEVRLN 747
 QY 200 FDGR-----G----- 204
 Db 748 SEGRELLDHSGEWINKESIIKDISKEYISFNPKENKITVKSKNLPSTLLOEIRNN 807
 QY 205 -----MVISNDTGT----- 214
 Db 808 SNSSDIELEEKVMLTECEINVSINSDTQIVEERIEBAKNLTSDSINYIKDEPKLIESISD 867
 QY 215 ---DVR-----HKAMRID---DDAKASMF-KKE----- 236
 Db 868 ALCDLKQONELEDHSFISFEDISETDEGFSIRFINKETGESIFVETEKTIPEYANHITE 927
 QY 237 ---DLKGT---DK-NYWLSDKI---PHAFNY-----YNGK----- 262
 Db 928 EISKIKGTFTVNGKLVKKVNLDTTHEVNTLNAAFFIQSLTEYNSSKESLSNLSVAMKV 987
 QY 263 -----IT-----VE---KYDDGRDYDP---HGMH-----IAGILAGNDTE 292
 Db 988 QVYAQLFSTGLNTITDAARKVVELSTALDETIDL-PTLSEGLPIATIIDGVSIGAAIK 1046
 QY 293 -----ODIKNFNGI-----DGIA-----P-- 306
 Db 1047 ELSETSDPLROEIEAKIGIMAVNLTTATTIITSSGLSAGSILLVPLAGISAGISPL 1106
 QY 307 -NAQ-----IFSY-----KMYSD----- 318
 Db 1107 VNNELVLRDKATKVVDYFKHVSIVETEGVFTLLDDKIMMPQDVLVISEIDFNNSIVLGK 1166
 QY 319 -----AGSGFAGDETMFAIEDSIKH-----NVDVVS----- 346
 Db 1167 CEIWRMEGGSG-----HTVTDIDHFSAPSITYREPHLSIYDVLVEVQKEELDLSKD 1218
 QY 347 -----SSGFT-----CTGLV-----GEKWO----- 362
 Db 1219 LMVLPNAPNRVFAWETGTPGLRSLENDGTLLDRIDNVEGEFYRYFAFTADALITTL 1278
 QY 363 -----AIR-----ALRKAGIPMVVAT-----GNVATSAAS----- 386
 Db 1279 KPREDYTNIRINDSNTRSFIVP-IITTEYIREKLSYFPGSGGTVALSQQVNMGINIE 1337

QY 387 ---SSSW-----DLV----- 393
 Db 1338 LSESDVMIIDVDNVVRDVTIESDKIKKGDLLIEGLISTLSIENKIILNSHEINFSGEVNG 1397
 QY 394 AN-----NHLK----- 399
 Db 1398 SNGFVSLTSLILEGINAIIEVDLLSKSYKLLISGELKILMNSNHIQOKIDYIGFNSQLQ 1457
 QY 400 -----MTD----- 402
 Db 1458 KNIPYSFVDSGKENGFGINGSTKEGLFVSELPDVLISKVYMDSDSKPSFGYYSNNLKDVK 1517
 QY 403 -----TG-----NVRTAAHEDAIASASA---KNO----- 424
 Db 1518 VITKDNVNLITGYLKKODIKISLSLTLODEKTIKLSNVHLDESVAETILKPMNRKGTNT 1577
 QY 425 -----TVEF----- 428
 Db 1578 SDLSMFLSMNIKSIFVNFLOSNIKPILDANFIISGTTISIGQFEFICDENDNIQPYFIK 1637
 QY 429 ----- 428
 Db 1638 FNTLETNYLYVGNRQNMIVEPNYDLDDSGDISSTVINFSOKYLYGIDSCVKNKVVISPN 1697
 QY 429 --DKVNI-----GGESF----- 438
 Db 1698 YTDENITPYETNTYPEVIVLDANYINEKINVININDLSIRYVNSDNGDNFILMSTSEE 1757
 QY 439 ---KYRNTGAFDKS---KITTN-EDGTPAKPSK---LKFV---YIGKGQDQLIGLDL 483
 Db 1758 NKVSQVKIRFVNFVKDITLANKLSFNFSKQDPVPSIILSFTPSYEDG---LIGYDL 1813
 QY 484 -----RGIAMVDRIY-----TKDLKNAFKKAMDK-----GA 510
 Db 1814 GLVSLYNEKFYNNFMVMVSGLIYINDSLYFYKPPVNNLITGFTVTGDDKYFNPINGGA 1873
 QY 511 RAI--MVVNTVNY--NR----- 524
 Db 1874 ASIGETIIDDKNYFNGSVGLQTGVFSTEDGFKYFAPANTLDENLEGEAIDFTGKLIIDE 1933
 QY 525 ---DN-----WTEL-----PAMG----- 534
 Db 1934 NIYYFDDNYRGAVEMKELDGEMHYFSPETGKAFGLAQIGDYKYFYNSDGMQKGFVSIN 1993
 QY 535 ---YEADEGFK---SQV-----FSISG-----DDGVKLWNNINPDKYTE--- 567
 Db 1994 DNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEE 2053
 QY 568 ---VKRNNK-----ED-----F-KDKLE-----QYYPID 587
 Db 2054 ISYSGILNPNKIYYFDDSFYAVGVWKKLEDDGSKYFYFDEDTAEAYIGLSLINDGQYYFND 2113
 QY 588 ---MES----- 591
 Db 2114 DGIMQGVFTINDKVFYFSDSGIIESGVONIDNVDYIDNGIYVQIGVFDTSQYKYPAP 2173
 QY 592 -NSNKP-----VG-----D-EKEID-FKAPDT----- 612
 Db 2174 ANTVDNITYGOAVEYSGLVYRGEDVYFGETYTIETGTWYIDMENESDKYFNPETKACK 2233
 QY 613 -----DKELY-KEDI--- 622
 Db 2234 GINLIDDIKYFDEKIMRTGLISPNNNYFENGENGMQFYINTDKMFFEGEDGVWQI 2293
 QY 623 ---VPAGS-----TSWGPRI DL-----LLKPD---VSAPGKNIKST 652
 Db 2294 GVENPTDGFKYFAHQNTLDENFEGESINYTGW---LDLDEKRYFYFTDEVIATGTSVI--- 2347
 QY 653 LNVINGKSTY 662
 Db 2348 ---IDGEBYY 2354

```

RESULT 7
AAR95011
ID AAR95011 standard; Protein; 2366 AA.
AC AAR95011;
XX
DT 08-JUL-1996 (first entry)
XX
DE C. difficile toxin B.
XX
KM Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;
XX diarrhoea; therapy; diagnosis; vaccine.
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
XX
PN MO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95MO-US13737.
XX
PR 07-JUN-1995; 95US-0480604.
PR 24-OCT-1994; 94US-0329154.
PR 16-MAR-1995; 95US-0405496.
PR 14-APR-1995; 95US-0422711.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
P1 Fitca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
P1 Williams JA;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29247.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin
PT useful to form anti-toxins against Clostridium botulinum type A, and
PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea
XX
PS Claim 36; Page 313-323; 434pp; English.
XX
CC Clostridium difficile VPI strain 10463 toxin B (AAR95011), the
CC product of the toxin B gene (AAT29247), is a cytotoxin associated
CC with diarrhoeic disease. It can be obtd. by expression in
CC transformed E. coli hosts of portions of DNA that together cover the
CC entire toxin B gene. Toxin B, and portions of it (see also AAR95012-13,
CC AAR95371-72 and AAR95018), pref. expressed as fusions to polypeptide
CC affinity tags or maltose binding protein, are used to raise avian
CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
XX
SQ Sequence 2366 AA;
Query Match 27.0%; Score 1188.3; DB 17; Length 2366;
Beet Local Similarity 12.4%; Pred. No. 4e-09;
Matches 292; Conservative 142; Mismatches 199; Indels 1717; Gaps 119;
QY 1 YP-----VV-----LADTSSSEDLNLSDEKVAENK-----SA 27
DB 34 YHNSSENTVVEKYLKDKINSLTDIY-IDTYKSGSRNALKKKFEYLVTEVLELKNNNLT 92
QY 28 --EKENHNI-----KKTAVI--KEKEV-----SKNP--VIDNNT----- 36
DB 93 PVEK--NHFVWIGQINDTAIYNQMKDVNSDYNVNFYDSNAFLINTLTKTYVESAI 150
QY 37 METSDPKE-----KKTAVI--KEKEV-----SKNP--VIDNNT----- 67
DB 151 NDTLESFRENLDPRFDYNNKFFPKRMEIITYDKQKFINYYKAQRENFELIIDIVKTYL 210
QY 68 SNEAK-IKENS-----NK-SQ----- 83
DB 211 SNEYSKEIDELNLTYSLESLKNTITONSQNDVNRNFEFFKGESEFNLYEOELVERMNLAAASD 270
QY 84 -----GDYTD----- 88

```

```

DB 271 ILRISALKEIGWYLDVDMLPGLQPDLPESIEKPSVTVDWMNTKLEAIKYEYIPEY 330
QY 89 -----SF-----VN----- 92
DB 331 TSEHFDMLDEBVOSSFVSLASKSDKSEIFSSLDGMEASPLEVXIAFNKGIINOGLISV 390
QY 93 -----KNTENPK--KE----- 102
DB 391 KDSYCSNLIYQIENRKYILNLSLNPALISEDNDNTTTTFIDSIMAANDNGRFMMEL 450
QY 103 -KVVYIAEFKDK-----SGEKA-----IKELS-----SKNTKV----- 131
DB 451 GXYLRVGFPPDVKTITNLSCGEAYAAAYQDLMFKESMNHLIADRLNREISKTNISQ 510
QY 132 -----LYTYD-----RI-FNCSALETTPDNLDKIQI-----EGISS 162
DB 511 STEQEMASLWSFDDARAKAOFEEYKRNYPFGSLGE--DDNLDPSQNIIVDKYLEKISS 568
QY 163 VERA----- 166
DB 569 LARSSERCYIHYIYQLOGDKISTEACNLFAKTYDSVLPQKNIEDSEIAYVPGGCI 628
QY 167 QKV-----QPM-----NMHARKE-----IG-----VEEAI DYLK----- 190
DB 629 QEIDKYKIPSIISDRPKIKLFFIGHGKDEFPTDIFAGDVSLSIIEAIDLAKEDISP 688
QY 191 -----SINA-----PFGK-----N 199
DB 689 KSIEINILGCMFSPYSINVEETYP-GKLLLVKDKKISLMPSSISQSIIVSANQYEVARIN 747
QY 200 FDGR-----G----- 204
DB 748 SEGRRELDHSGEWINKESIIKDISKEYISFNPKENKITVSKNLPSTLLQEIERN 807
QY 205 -----WVISNIDTGT----- 214
DB 808 SNSSDIELEKVMLTCEGINVISNIDQIYVERIEAKNLSDSINVIKDFKLEISID 867
QY 215 --DYR-----HKAMRID--DDAKASNR--KKE----- 236
DB 868 ALCDKQONLEBDEHPSFEDISSETDEGFSIRFYNKETGESIFVETEKTIPEYANHTE 927
QY 237 --DLKGT--DK--RYWLSDKI-----PHAFNV-----YNGK----- 262
DB 928 EISKIGTIFDTYNGKLVKKVNLDTTREVNLNAAFTQSLIEYNSKESLSNVAMKV 987
QY 263 -----IT-----VE-----KYDDGRDYFDP--HGME-----IAGILAGNDTE 292
DB 988 QVNAQLPSTGINTITTDAAKVVELVSTALDETIDLL-PTLSGCLPIATIIDGVSLGAIR 1046
QY 293 -----QDIKNFNGI-----DGI A-----P-- 306
DB 1047 ELSETSPDLLRQEIIEAKIGIWAUVLTATTAIITSSIGISGFSILLPLAGISAGIPSL 1106
QY 307 -NAQ-----IT-----ISY-----KMSYD----- 318
DB 1107 VNNELVLDRKATYKVVDYFKHVSIVETEGVFLLDDKIMPODDIVISEIDFNNSIYLGK 1166
QY 319 -----AGSGFADETWIFHAIEDSIKH-----NVDVVSV----- 346
DB 1167 CEIWRMEGGS-----HTVTDDIDHFSAPSITREPHLSIYDVLEVOKEBELDSKD 1218
QY 347 -----SGGFT-----GTGLV-----GEKYWQ----- 362
DB 1219 LMTVLPNAPNRVAFWETGTPGLRLSLENDGTCLDRIRDNVEGEPYWRFAFIADALITTL 1278
QY 363 -----AIR-----AIRKAGIIPWVAT-----GNVATSA----- 386
DB 1279 KPRYEDTNIRINDSNTSRSPITV-IITTEYIREKLSYSPYSGGTYALSQYNNMGINIE 1337
QY 387 --SSSW-----DLV----- 393

```



```

QY 97 N-----P-----KXEDKV-----YI 107
DB 265 NNRHSVONIKPHSGSVLNRETLGCTQSAYGRVPIVRGNHSGKLYDARIKLSQIRRKI 324
QY 108 AEFKD-----KESGE----- 117
DB 325 VRFADPGVFMNGFNTAFQRYRKQDREHTCETDLDBEEGCEVAALLCLALPFGKITCNK 384
QY 118 -----KAIK 121
DB 385 CVEENLSEGOATHENILKKQEIIRHIVLQRHPQKHALQILERSKALQSVNSNYKQFT 444
QY 122 ELSLKNKTVLYTYR-----IFNGSAI-----ETT----- 147
DB 445 EIHSLSECKTLPAFQOARINDVLKGGSATAEISEATRNILBEIVRYLKARTESSEKGT 504
QY 148 -----P-----DN-LDK-----IKQIESISVE 164
DB 505 LKTFNKTSQKALNPALMCDNQDENGNFVWGERGYHAKRFENKYFELVDBSKGYAKPE 564
QY 165 -----RAQ----- 167
DB 565 ARINRGORKTAITRLIVPTNEVLRQEQGESIGENPLTVBCTSVLNGDFLPPCCVYN 624
QY 168 -----KVQ-PKMNH----- 177
DB 625 EAGEPISELQMPKQHLVNGSGSKYVDMPPQEGQMYIAKAGCYWNIPLANLVNR 684
QY 178 KE-----IGVEE----- 184
DB 685 KEEAKATKQWADVLIINGTWPILLDVASACYLLKVFPPDSASAEPRIMVDHKTMTM 744
QY 185 AIDYKXINA-----PF--GKNPD-----GRGVISN--IDTGDYR-- 217
DB 745 VVDSGSLNTGYHILKANTVEQLIFTRAGLSDMKHYLVG-GPILNEDIDP-TEYRTP 802
QY 218 -H-----KAMRID-----DDA 227
DB 803 SMLRLKGIYNPVOLLDIRIDIRLYLPLYALLSPGVLIAMNSASLEILREYLRKDE 862
QY 228 KAS----- 230
DB 863 FVSIVLILESLARKVSVSTLSQMLIEGEAQYIIIEAVQIKQRYPIPTYVMMEMLIL 922
QY 231 -----MREKEDLKGTQNY-----WLSQKI PHAF----- 255
DB 923 ASRSESDALDAAGFKKFORRESIQLMEXKYRLILEDEWRELSLRORFSATILRSSKFAMRT 982
QY 256 -----NY-----NG-----GKIT----- 264
DB 983 HGGELNASI EDIGRYSSESMNYFQELKQVMIKYKITNOAKVITOSTHTSIKRYVSC 1042
QY 265 -----VEKYD-----GRDY----- 274
DB 1043 FNYLIPDVSKFINWVCLMILTLMQELHTWERTNRCKRIARRENEQEKHKIKPMQA 1102
QY 275 F-----DP-----HGMH-----IAGILAG 288
DB 1103 FQNEHKVDPTEFEFLEYLQKHTPELLTYFOEDEVVVHQAQRGSELEBRVAFIALVMV 1162
QY 289 NDTQD-D--IKNFN-----GI-----DGI--APN 307
DB 1163 FDSERSDCVVKILNKLIISSTDAVYHQGLSEIEDDINEKNLITIDFELSODGVTRDPN 1222
QY 308 A-----QI-----FSYK----- 314
DB 1223 VMEHTFSTMWMTQVONGRTIPIHYRTBEGHFMTRANAHVAITEITNEHKDMLMGAVGS 1282
QY 315 -----MYS DAG----- 320
DB 1283 GKSTGLPFLHSKRGVLLVEPTRPLAENVYRQLSHEPFIYATLIMRGLTTCGSSPVTIM 1342

```

```

QY 321 -SGFA----- 326
DB 1343 TSGFALNQLANNRRIAEYDFVIFDECHVDANAMALRCLLHDAEFPGKVIKVSATPPGR 1402
QY 327 ETMF--HAIE-----DSIKH-----N--VDVVS----- 345
DB 1403 EVEFTQHPVLLTEETLGLKEFVDAQGTGVNCVIRHGDNLVLYVASCNEVDIISKALI 1462
QY 346 -----VSSG-----FT-----GTGLV-- 356
DB 1463 DKGHKVTQDRTKWKGVKVEIITSGTPQKRFVVAATNIENGVTLDIEVVVDFGTVVPF 1522
QY 357 -----GEKYQAI RALR--KAG-----IPMVA----- 377
DB 1523 LDVDRMMQYOKVAINGERIQRLGRVGRHKAQTALRIHTERGLSEVPSCIALTEAARFC 1582
QY 378 -----TGNVATS----- 384
DB 1583 FTFGLPVITNNVTLSLSNATVRQARTWAFELSPFYTHFVRVYDGTWHPHVKLRFK 1642
QY 385 -----ASSSS-----WDL 392
DB 1643 LRDBEIVLNTKTAIPNRGVNTWMTSSAYORLGANVDSNEIRIPLCKEVPETLHETIWDI 1702
QY 393 VANNHLKMTDTG----- 404
DB 1703 I-TTH--KSDAGFGRLSASACVAYTLKTDVMSIORTIHIIIDALIYEBROKOEYFRIT 1759
QY 405 -----NTR--TAAH--EDAIVASAKQTVBPKVI----- 433
DB 1760 TNSISSNFSQSIANAIARFSSDHTVENISVLNKAQCEFPNLIIDAAPODPOSOV 1819
QY 434 -----G-----GESFKR-----NIGA----- 445
DB 1820 GRSYISNEGALDAVYHOSSEKAMSEHFKLKGWNPILTRDIMIGVLLGLMMVYQOK 1879
QY 446 -----FPD--KSKI--TTNEBGTK----- 460
DB 1880 SQMTEVYHNEAKGKQORORLFRDARBSKLGREYVGGDGTMEHYFGAETYKKGKTSKXK 1939
QY 461 -----AP--SKLKFV----- 468
DB 1940 GMGOKQRFVVMYSFDPEDFAVRFDVLTGATLDETPTIDLHLVQEHFTKIRSEMBESG 1999
QY 469 -----YIGK----- 473
DB 2000 ELBSQHLVSGVAVAYNNNRGTALQVDTLRPNPLVLCANKPTIAGPEREYELRQTQ 2059
QY 474 -----OD-----OD-LIGLDR 484
DB 2060 PKAISLKDVPKANDLSEKVOHESASLHRLGRDVPNIPISNICKLINRSEGERDVTWGLGF 2118
QY 485 GKIAVDR-----TY-----TKDLK----- 499
DB 2119 GPVITTRNHLPEHNGGELDITRHHGDFLITMTKLOLYPVNRRDLILRLPKDIPRPFQK 2178
QY 500 -----NAFKKA-----MDK-----G 509
DB 2179 LQFROPERNEKICMVGNSFOAKSVNTVSETSILPMDCHCFWKGWITTKDQGGCGLPVS 2238
QY 510 ARAIMVY-----NTVNY-----YNRDN----- 526
DB 2239 TRDGNIVGHSLSGFNNITINFASF PENFVNQVLYLTPEBNHOMIOMWKYNTDNISGALKI 2298
QY 527 -----W--TEL----- 530
DB 2299 SNEAPGLFKTKLIGDLESIFVROQMKREKMYVTJOLDGNKAIASCPNOLVSKYVVGK 2358
QY 531 -----PAMG----- 534
DB 2359 CPMFDMYLKLEAGRKFTPLGQYOKSRUNKKAVYIDIMKYSTVIEAGNVQOTPEDAV 2418
QY 535 -----YEAD-----GTSQGVF-----SI--- 548

```

Db 2419 RLLIQLSELGFCQVITDEDFVFNALNMSAVGALYGGKKDYFKDFTQEMKETILKQ 2478
Qy 549 -----SGDDGVKLWN-----MINPDKTEVVRNKK-----EDPK- 577
Db 2479 SCARLYTGKG--LWNGSLKAELRPLEKV---QANKTRFTTAAPLDTLLGGKACVDDFFN 2533
Qy 578 ----- 577
Db 2534 QFYELNKGPNVGMKPYGGMNELLTKLPDGIHCDADGSDQFSSLSPYLINAVNLRL 2593
Qy 578 -----DKLEQ-----YYPI---D---MESF---NSNKP-----NV----- 598
Db 2594 HFMTWIDIGQMLNRLYTEIVYTIPIATPDGTIVKKFGNNSGQSTVVVDNTLMVLALKY 2653
Qy 599 -----GDEKE-----ID-----PK-----FAPDT-DK 614
Db 2654 SLLKDGVEAEKHKQVIKVFVNGDILLISIDPAYEGLLDTMQGNFKELGKYDFNSRTRDK 2713
Qy 615 -ELYK-----EDIIVP-----AGTSMG----- 631
Db 2714 GELFMHQGRKVEDIWIPLKEQRIVSILEWDRSKPEGNRMEICAAMIESWGHOELTH 2773
Qy 632 -----PRI-----DLILK 639
Db 2774 QIRFYAWLIGQAPYSGLAETKAPYIAESALRKLKLYLDKDQSAIEVYLRAIFEDYITE 2833
Qy 640 P-----D-VSAP-----GK----- 647
Db 2834 PEDLVVYHQSGDLDAGTSAPSKARKQESASSTQGIAPTMEGPEPADDPGKSKQQTVI 2893
Qy 648 NIKSTLV----- 655
Db 2894 NLEKDVNVGTGTFAVPRLKGLATKMSMPRVGKAAMNLDHLLVYNPQVDLANTRATRK 2953
Qy 656 -----ING----- 658
Db 2954 QFDTWYGVKRDYELDDSSMQIILNGLMWCIENGTSNPNMGVMMMDGEEQIEVPKPL 3013
Qy 659 ----KST-----YGY 664
Db 3014 IDHAKPTFRQIMAHFSYVAEAYIEKRNOEKAYMPRYGL 3051

RESULT 9
ID ABG23329
XX ABG23329 standard; Protein; 6619 AA.
AC ABG23329;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #23320.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX OS-
XX WO200175067-A2.
XX PN
XX PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX

DR N-PSDB; AAS87516.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 53688; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6619 AA;

Query Match 26.7%; Score 1175.3; DB 22; Length 6619;
Best Local Similarity 9.4%; Pred. No. 3.8e-07;
Matches 345; Conservative 151; Mismatches 142; Indels 3047; Gaps 153;
Qy 2 PV-----VLAD-----TSSSEDALNISKD-----E 21
Db 2746 PVKATDVKPIKKTWEILKDLWNQFLKAEISNEAITLADEIIAAKGCKKSWASLFGSGD 2805
Qy 22 KVAE-----NKEK----- 29
Db 2806 MVQKLIDARANKPKPEKAVKODKTVAPKKTVTTRSNADVIRSVLNDKMGSDYCKAHAIH 2865
Qy 30 -----HENIHAME----- 38
Db 2866 DVKRAAKSYAALGLTPCECSQIDKLKHLKIHTMAKYVRANKRAHDDDDIIRMNVAVLGL 2925
Qy 39 -----TSQDFKE----- 45
Db 2926 GTIAKTLGVHPTVTTLRLSLGIEPADTRRTFMENVLAPLPTHVADWLSEQVGPKYEIRQ 2985
Qy 46 -----KKTA-----VIK----- 52
Db 2986 YVRDLLLEAYNNRHNHKTAEHERFLKRYVEVVEMLEELAPNSEAYDTLHNAIAALTELS 3045
Qy 53 -----EKEVVSKN-----P-----VIDNNTS----- 68
Db 3046 ELMKSSDKVEYAPNSNRLAILDALADQIVTATGIGTFLGMNVPGALTEVNRSNYSKFDG 3105
Qy 69 -----NEEAKIKEE----- 77
Db 3106 EPVFNENKYNKKEAAAEAGFFDPLDPNASELRISGPGTGKTFTHAHMIDEIMPRYHETC 3165
Qy 78 -----NSNKS-----QG-----DYTDS-- 89
Db 3166 SLMGIPALYNEVIMTATTNKAEEVLARATGRPTSTYHSFQGLTVRNNLTKTGADLIIPSKS 3225
Qy 90 -----FVNKNTENPKKEDKVY----- 106
Db 3226 FSIKKNKVIFIDEASMDIRLKKKFINEGTH-----QSKIVYLGDKCOLLPVKETSSPVYND 3281

QY 107 --IAEF-----KXSEGEKAIK 121
 Db 3282 SSIKEFMLTQOMRTDVPBELHALHEQURGTIEGKHGFLPIKAVPGIIMDKD--GDEMRK 3338
 QY 122 ELSSLKATK---VLYT-----YD--RIFNG-----SAIET--TPDND 152
 Db 3339 EVEGHFMKTDRIVAYTNQOVVNYNFTIRANNGHGECEICEBELVNSAIOIGADR- 3397
 QY 153 KIKO-----IEGI-----160
 Db 3398 SIEQECQIVDODSATRKILIDSSGIELEVEDCTLDTGCGVEGEPIDVMDYFIRLOKX 3457
 QY 161 -----SSVERAOK-----VOP-----171
 Db 3458 YAHQKWBHRYLKEFPBELRALHASTVHKSGSSTYDTIFIDATDLSRCQPDVAVRLY 3517
 QY 172 ---MMNHR-----KEIGVEBAI-----D-YLK 190
 Db 3518 VAVSRARH-RVVFYGSIAEKFGELMOLG-EKOIREMTNSELVQELFLABELRLDKYLD 3575
 QY 191 SINAPFGKNF-----DGRGVISN--ID-----211
 Db 3576 SITNGETFAVYNSTGFLYLCEFYSSRSQOPSGHRLALANHLDDMREYLEQAGL 3635
 QY 212 -----TGTDR-----H-KAMR-----IDDDAKA 229
 Db 3636 LLEVHVMQVPRVLQGCCTYQDVARDALBECILANDKQCHLSLQRTKRAATLNDMA 3695
 QY 230 -----SMR-----FKKE-----DLKGTD--KNY--WLS 248
 Db 3696 LRHMRSGLSNTMBEPILSFEVCMRYITFNNEGIYPALITTLTKRCDIVAVAMEPMDV 3755
 QY 249 DK-----IPHA-----NY-----258
 Db 3756 DDOSMILELHATPAKGKKTPAKMOEFTTQELVP-AFEAQTQOYIICGSEYVQOLTKOG 3814
 QY 259 ---NGC-----KIT-----VEK--267
 Db 3815 KADANIGVMDCAYNOKIYVPNRYQIFYDPDKIKTKIALSMQALIDVSGSYVEPGIK 3874
 QY 268 -----YDDG-----271
 Db 3875 IHHVADYLPBPAISKMLDLEMDVPLAVDIEFSLKHVDCGIGTIFCWNKHEGIAFP 3934
 QY 272 -----RBYF-----275
 Db 3935 VDYEPIEGATEAPRGQVNNMLVRSMLRDFIKYLRQWYHNIAPDYVALIYQLEMTDL 3994
 QY 276 DP---HGMI-----AGI-----LAGN-----289
 Db 3995 DTEGLHGHMSIMLRWMDCTKLITYLATNSCAGNKLSLKDQAOEYAGNYAOEINDITRP 4054
 QY 290 -----DTE-----292
 Db 4055 LAELLEYNLVDGLCTWYVEYKEMDTLVNDQOLDVYTNIFKPACEIDIIQOLTGMPINRT 4114
 QY 293 -----QDIKN-----FN--299
 Db 4115 VLEVEALTDYNNALKTIAISKVJIDFTRLNBEWEKONQILKKKRVTLADAKEQENP 4174
 QY 300 -----GI-----301
 Db 4175 NSGIQOLKLEFFLGULPVGLTASKLPRATSGILSLKKNHTODTSLBELDALIYKAVD 4234
 QY 302 -----DG-----IAPNAQ-----309
 Db 4235 KILTAIPALKVARQCPDGMHYLFGNLNGTIVSGRLSSSEBNLQNLPSGSRYAKMIKC 4294
 QY 310 ---F-----SYKMSD-----318
 Db 4295 FEAPGWIFCGLDFASLEDRIISALTTKDPNKLKVTYTDGVDGSHLRAVAVFGEMPDIEDT 4354
 QY 319 -----AGS-----GF-----323

Db 4355 VESVNSIQEKYKAYRODSKAPTFALTYQGTITLKNKGCFPEQKABVEERYHTLYKUSD 4414
 QY 324 -----323
 Db 4415 DMVQAKLDQAAKDGVTVAFLRVTRPLLQOVRIGTSKTPYEABEGRTAGNALGOSWCL 4474
 QY 324 ---AGDE-----TMFHA-----IED-----335
 Db 4475 LNNRAGSEFMRKVRNSKRRLDIRPSTHIDAOYFLIRDDMDVYITNTHLVYKAVQWODHP 4534
 QY 336 SIKH-----339
 Db 4535 DIAHPDVHLDDQEGAVIVQNTMLLTBEKQVTFRDLGRAQSAALPQNLNDRFGQVVDVRD 4594
 QY 340 ---NV-----DVV-----344
 Db 4595 FLNYSYLGHMSQPDFOQNLVTNPQRMQOPRYOABCEIDIQALNMTFDEGCLFKALMRNA 4654
 QY 345 -----SVSGCF 350
 Db 4655 ARLNGKFGNNAVYDAEMHYANRIFIKESGQKKAEBEPVSALKKGBEIHMSSELFL 4714
 QY 351 T-----351
 Db 4715 TLLTGEAAGEVMSLAGHHTVPRRLGVMPIAVAMSKAEQVDVVRADGTTITHVPRHEID 4774
 QY 352 -----351
 Db 4775 WSDKIAPPARITAMRFVWMLHDEYDVKDKYLSVTLPLKPIKHIVMKHRVDSVESMD 4834
 QY 352 ---GGL---VEGKYQOAR--ALRKAG-----IPM 374
 Db 4835 IMDVSTMGTLHDSIEKAWNHGHNALRLKLGPERVVQVNVINPTPEQIASPNLIPV 4894
 QY 375 ---VATG-----379
 Db 4895 WIEQATKIKGWTIGCFDITVEGLLDQVSTSTYTVKSGRDEBENHLOQSMYRWLHDD 4954
 QY 380 ---NY-----ATSA-----SSSW--390
 Db 4955 KITEVDIRINYIFTDMMKALATSOENYKHLKHDIPMSYBATEEWIKAKLALIDRYW 5014
 QY 391 -----390
 Db 5015 DAPSEIPECTDEELWRSDPVFKYFADBEKAKQOPASTKNPPTLFEARKFMAEKGCTI 5074
 QY 391 ---DLVANNHLK-----TDTG-----404
 Db 5075 LHVEPIFPMIDLTVGSHHPALBEIYDVLCNKTOITDRGFPFVEVAYFLAKMA SCGATII 5134
 QY 405 ---NTRTA-----AH-----EDAI-AVA-----419
 Db 5135 TKDRGEIPVNIYAVVALATSGFGKHSVNIVEDGFMAFGFKRFMEDTMBALBRLMKIAN 5194
 QY 420 ---SAKN---QVEEDKVNIGESFKRYNIGAF--FD-----448
 Db 5195 ERSARNATDOKEBDKV---EA-EYRTGAYPFTFSGTPRAYKOLYHNKLLAGCGAIN 5249
 QY 449 -----KSKITTN-----E-DG--TKA-----P 462
 Db 5250 LOIDEIGSNLLANADVLTVPLELYDQGIKKOKLTKNTAESIRSEBVDKTFANLLPSTP 5309
 QY 463 SKL-----KFVY-----IGKG-----OD-----475
 Db 5310 SKLINGSQTEELFPYDLDTGYSRRCLFVAGQADKRAVLSGABEYRNLIKODNNAAVVK 5369
 QY 476 -----ODLIG-----480
 Db 5370 WANHFHSLADSKFGKGMKTVEDDVGIALIDYKIOCEKQAAALADHEIRKAELEHRYFKA 5429
 QY 481 LDLRGKIAVMD-----RIYTK-----496

Db 5430 LKAGAFAPVDSINVMEHLKQAILLVEESGVAFTILNREKAVVYKLAKYIASVGTVEV 5489
Qy 497 --DL-----KVA-----FK-----KAMD----- 507
Db 5490 HADLLEALPFYKSGNAERNMTLATAMGYKQHIIKKTYNGBIEFFRGELTKETDTNEM 5549
Qy 508 -----KGARA-----IMVVTWN 520
Db 5550 IVSYGENFAYDIABKVPDPQLHLVLTQOAGYHAWNHFRGRHRAENALAGNMMIVD 5609
Qy 521 -----Y-----NR-----DN---WT 528
Db 5610 GTCSLDVCHLMKEYRFYTTTKRHTDEENRRLIIPMNYQLELDGAEYKEFMDNMVMAW- 5668
Qy 529 ELPANGYRADE-----GT-----KSQVF-----SIS 549
Db 5669 -LP---FKTDESANQRAKKWESCAKGTYYNNDAFLDVRDFIPRTSKNEAFQOQMKVQ 5724
Qy 550 GDDGVKLM-----N-MI-----NP---DK-- 564
Db 5725 NLDNLERMPAORIATGNRNHMIKVALALVDSGIGFNEVKQVRHEFNKLSNPLSADELD 5784
Qy 565 ----- 564
Db 5785 STVMITVAKLLFLVBEDCOLKEETWSEBEFHGEMNDQLILIVGSAAGKSASLRNROE 5844
Qy 565 -----KTEV-KR-----NNK-----BDF-----KDKLE 581
Db 5845 RVVYLNTAGKRLPKFNNGVRITDPYQILEYFDQCIEHRDDVDGIIVDSLTFMMDMLE 5904
Qy 582 -QY-----PI-----DMESEF----- 591
Db 5905 TOYVLTANTQKAMGEFAOFFKILLQEKVKVFAKPVIFTAHVKDEVDERAMELKTVPVK 5964
Qy 592 -----N-----SN-----KP----- 596
Db 5965 GSKKNGIBAYFSTVSAERIDLKEKYSNGMLEITDEQELGYKHVFQTRPKTKTVCK 6024
Qy 597 -----NVG----- 599
Db 6025 RIRSPMGFESKOETYMNDNAQKLLDHLAEYYSTGOAAKAEKATDNLGGFGAKESDIYLA 6084
Qy 600 ----- 599
Db 6085 NLKAYAGKAASGANFIQIIADLTLDHGSAGEYREQLYITSGTEKCKCTYEKNGKEYF 6144
Qy 600 -----D-----EKEI-----DF-----KFA----- 609
Db 6145 LPGYTVINDILVMTSGETIPEAVFEKVVNVYDFDEKKEVAKSVMPVNAIGKFAVAIL 6204
Qy 610 -----PDT-----DKELYKEDIIV- 623
Db 6205 KSEEDKQTKDGSNGVYVSTGETRFTNTIEKVFHPDLHLTVVEABELTGERKELTVEEAVFM 6264
Qy 624 -----P-----AGSTS----- 629
Db 6265 DKWLEKNKGVTRDKTKTGASGAKAGOPPKPGATNTGAGASAAKSLPEIMQIIVLGVDSRF 6324
Qy 630 --WG-PR-----IDLL-----KPDVSAPG-----KNIKS-----TLNVI--- 656
Db 6325 RNWGLARGMLDLETGLSLGLDLKLVTETPD---GTKQVRQSKQMQAAEDITTGVIDWF 6380
Qy 657 -----NKSSTVGY 664
Db 6381 KEAKVIFVEVPVQSQANGMKSYGV 6405

RESULT 10
ABG23295
ID ABG23295 standard; Protein; 5909 AA.
XX
AC ABG23295;
XX

DT 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #23286.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
XX N-PSDB; AAS87482.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 53654; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 5909 AA;

Query Match 26.7%; Score 1173; DB 22; Length 5909;
Best Local Similarity 9.2%; Pred.No. 2.7e-07;
Matches 311; Conservative 169; Mismatches 153; Indels 2730; Gaps 130;
Qy 1 YPVVLADTS-----SSDALN-----ISD-----KE----- 21
Db 2305 YAVWITVDPRNHRIVHVIRESGRTSEYVNNLLAQTPFKVFDLWTQAKEFVRL 2364
Qy 22 -----KVAE--NKEK-----H----- 30
Db 2365 NNTQGVISVYNKLEDGIPVKVAVKATEPTNKEKLEAASTKLKMGAVATDAYEFTDLNE 2424
Qy 31 -----ENIH-----SA-----METSQP-----F 43
Db 2425 NEETIVAVNDKVPDQVQNLHRRHQSANLKDYGFTKFLERLSTVIKRRHSVEDLMKF 2484
Qy 44 KE----- 45

Db 2485 MEKGDPLIADGSIYFKRLKESKDNPEHLKQVMDVDFVDCSHKIEQCVKRWKENLV 2544
 QY 46 ----- 45
 Db 2545 DQNRKDCSHGLHVASLOYIRNFGSVNTTIGKVADEVFVAPEYDVTKRWASAHITKL 2604
 QY 46 ----- 45
 Db 2605 PDALFDHVNNGNPITIEGCTELMMVLSGNHPERSQOVUGGHYGNLTYSELTRNYLD 2664
 QY 46 ----- 65
 Db 2665 DDEDSVIPDEPIVKKKTALNMESFATNPKAEPKATDVPIKTKTMEILKQLMNQFLKA 2724
 QY 66 NTSNE-----EAKIKEENS-----NK----- 81
 Db 2725 ETSNEAITLADBEIIAAKGCKCKSMASLGFGDMQKLI DARANKKRPKAVKQDVTAPK 2784
 QY 82 ----- 85
 Db 2785 KTTVRSRNVADYRSYLDKMSDYCKAAIHIDVKAACKSYAALGLTPECSQIDKLKH 2844
 QY 86 ----- 95
 Db 2845 HLKHTMAKVYRANKRAHDDIIIRMAVAGLSGTIAKLTGVHTVTLRLSGLIEPADT 2904
 QY 96 -----EN-----PKKE----- 101
 Db 2905 RRTFEMENYLAPLPTHVADWLSEQVGPKEIRQYVBDLLEAVNNRHNKETAHERFLKRY 2964
 QY 102 ----- 107
 Db 2965 VEUVMELEBANSEAYDTLHNAIALTELSELKSSDKETVAANSRNLALDLADQI 3024
 QY 108 ----- 124
 Db 3025 VTATGTFGLMNVPGALTEVNRSVNSKDEGEPAFENENKMYKMAABEGFFDLDPNA 3084
 QY 125 ----- 133
 Db 3085 SELRISGCGTGCTFTMAHMIDEIMPRYHETCSLGPALYNEVIMTATNKAEBVLARA 3144
 QY 134 ----- 137
 Db 3145 TGRPTSTHSGOGLTVRNNLKTGBADLIPSKSFSTIKANKVIFIDEASMDRKLKFFINEG 3204
 QY 138 ----- 143
 Db 3205 THQSKIYVLGDKCQQLPVKETSSPYVNDSSIKERMLTQOMRTDVPHELALHEQLAGTIBG 3264
 QY 144 ----- 174
 Db 3265 KHGFLPIKAVPGIIMIKGDDEMRKEVEGHFMTKTDTRIYATNOQVNVNTPFIRAMNGF 3324
 QY 175 HARKSIGVEE-----AIDYLK----- 193
 Db 3325 HGEFCTIG-BELVSNASAIQIAGDRLSIEOEQI VDOBSATRKILIDSSGIELEVRDCTLD 3383
 QY 194 APFGKMPDG-----RGMVISNI-----DT-----GTDY----- 216
 Db 3384 TGYGVFPFGIPIPKPEFPFLRALHASTYHKSOGSTYDTIFIDATDTLSTCROQDVVARLLY 3443
 QY 217 -----RHKA----- 222
 Db 3444 VAVSRARHVVYFYGSLAEKFGGEFYSSRSQOPSGHRLALANHLDDMBEYLEQAGKL 3503
 QY 223 ----- 229
 Db 3504 LLEVHVNOMVFRLYOGCKTYVDVBDALPECLIAMDKQGLKSLORTKKAFTLLDNMA 3563
 QY 230 -----SMR-----FK-----KEDLKGTDKRWYLS 248

Db 3564 LRHNRKSCRLSNTWREPILSFEVCMRYITFNNEGIYPIALITTLRLKQDVIKAVMEPMDV 3623
 QY 249 DK-----IDHAF-----NYU----- 258
 Db 3624 DKDSVMIELHTAPGKKTPAKEMOEFTIOELVP-AEPAQTOYIICGDSERYKQLTQOG 3682
 QY 259 -----NGC-----KIT-----VEK----- 267
 Db 3683 KADANIGYVMDCAVGNQKVIYVPRYROI FYDPDKIKTKIALSMQALIDYSGSYVEPQTK 3742
 QY 268 -----YDQG----- 271
 Db 3743 IHHADYPLTPEAISKWLKLEMDVPLAVDIETFSLKHYDCGIGITIFCWNKHGIAFP 3802
 QY 272 -----RDYF----- 275
 Db 3803 VDYEPIEGATEAPYGRQVHNMLVRSMLRDFPIKYLNRQWYHIAFDVYALIYOLFMTDL 3862
 QY 276 DP-----HGMHI-----AGI-----LAGN----- 289
 Db 3863 DTBGLHGMSTMLANWDCSTKLITLATNSCAGNLSLKDOAOBYAGNVAOEINDITRIP 3922
 QY 290 -----DTE----- 292
 Db 3923 LAELLEYNLVDGLCTWYVEYKHWDTLVNDQODVYTNIPKPACEDIIOMQLTGMPIINRT 3982
 QY 293 -----ODIKX-----PNG 300
 Db 3983 VLEVEEALTDYNNALKTIAISKVYI KQFTRLLEBWEKONQILKKKVTIADAKEQPNP 4042
 QY 301 IDGIA----- 305
 Db 4043 NSGIQLOKLFEPGLPVLGSLASKLPATSGILSKLNHTQDSILEILDALIDYKAVD 4102
 QY 306 -----PNQO----- 309
 Db 4103 KILTAPIALKARQRPDGMHYLFQNLNLGTVSGRLSSSEPNQLNPSGSRYAKMIKC 4162
 QY 310 -----IF-----SYKYSAGSGFADETMFHA----- 332
 Db 4163 FEAPPGWIFCGLDPASLEDRIASALTTPDKPNLKVYTD---GYDG---HSLRAYAVPGE 4214
 QY 333 -----IEDSIKHNVSYSV-----GF---TGSG 354
 Db 4215 MNPDIETVE-----SVNSIQEKYKAYRQDSKAPTFALTYIQGYITLTKMKXCGFPEQOAR 4268
 QY 355 LVGERKY-----W-----QAIRALKRAGIPMYV 376
 Db 4269 MVEERYHTLVKSDDWQAKLDQAKADGYVTAVGLRVTRPDLQOVIRGSTK---PYEAB 4326
 QY 377 -----ATGNVATSSASSW----- 390
 Db 4327 AEGRTAGN-----ALGQSWCLNNRAGSEFMKRVNSKRLDIRPSTHIDAQYFLIRDDM 4382
 QY 391 DLV---ANNHLKMT---DTGNVTRTAHED---AIVASAK-----NOT 425
 Db 4383 DVVITYNTHLVKAVQWQDHPDIAPBDVHLDDQEBQAVIYQNTMLLTKEQVTFRDLGRAOS 4442
 QY 426 VEPDKVN-----IG----- 434
 Db 4443 ALFQNLNDRFGQVVDVDRDIVFLVNSYLGHSQRPQOQLVTPQRMEOQPYOABECEDIIQ 4502
 QY 435 -----GESFK---YRN---IG-----AFPDKSK---ITTNEQG---TKA 461
 Db 4503 ALNMTFDEGCUFKALMRVAAARLNGKRGUNAAVYDAEPMVYANRIFIKESGGQKEAE 4562
 QY 462 P-----SKLKFVYIG---KGODODLIG----- 480
 Db 4563 PVSAAKKGEBEIHMSLSBELFTLLTGEAABGEVMSLAGHHTHVPRNLGMPPLAVAMSKAEQ 4622
 QY 481 -----LDLRGKIA-----VMDR1Y----- 494
 Db 4623 VDVVRADQTTITTHVFPHEIDMSDKIAPARITAMRFTYWLHDEYDVYKDKKYSVTLL 4682

Qy 495 -----TKDLKNAFKKAMDKG-----ARAI- 513
 Db 4683 KPIKIHVMKHRVDSVESMDIMDLVSTSMGTGLHDSIEKAWHGHNAALRKLGYPERVVQ 4742
 Qy 514 -MVNT-----
 Db 4743 NVVINTPEQIASNPILIPWIBORATKGIAGTWIGGKFDIVTEGLLODVKSTSTYTWVK 4802
 Qy 519 -----VNYNRDNMT-----LPA 532
 Db 4803 GGRDEHRLQSGMYRWLHDDKITEDVIRINIFTD-WMKALATSQENYPKRVHLKDIPL 4861
 Qy 533 MGYEA-----DE-----539
 Db 4862 MSYEATEWIKAKLALIDKYWDAPSEIPECTDELMRSDPVFKYFADPEKAKQPGARST 4921
 Qy 540 -----GT-----KSO-----V 545
 Db 4922 KNFTLFEARFKMAEKGKGTILHVPEIFPMIDLTGCVSHHPAIEIEIVDVLCNKTQNTDRGF 4981
 Qy 546 FSI-----SG-----DDGVK-----555
 Db 4982 FRVEVAYFLAKWASCMGATIITKORGEIPVNIYAVALATSGFGKHSVNIVEDGFMAGFK 5041
 Qy 556 -----LNMNIPDKKTEVKRNN-----KEDFKDLK-XY-----YPIDMESF 591
 Db 5042 KRFMEDTMPALAEADRLMKIANNERS-----ARNATQKEEF-DKVEAEYRRTGAYPF---TF 5093
 Qy 592 NSNKP-----N-----VG-----DE-----601
 Db 5094 DSGTPPAVKQRHLKLLAGCGAINLQIDEISNLLANADVLTVFLELYDQSKIQKLTKN 5153
 Qy 602 -----KEID-----FKF-AP-----DT-----DK-----614
 Db 5154 TAESIRSEVGDKTPANLFLGTSPKLLNGSQTELFYDFLDGTGYSRRLFAVGQADKRA 5213
 Qy 615 -----ELYK-----ED-----620
 Db 5214 YLSOSAIEIYRNLIKQDNNVNVKWNHFLSHLADASKFGWKVTEDDVGIALIDYKIQCE 5273
 Qy 621 -----IIV-----623
 Db 5274 KQAAALADHEIRKAELEHRYFKALKAGAFVDSNIVWEMHLKQAILLVESGVAFQ 5333
 Qy 624 -----P-----AGSTS-----W 630
 Db 5334 TILNREKAYVKLAKVIASVGTEVTHADLLEALPFYKSGNAAQKVPFDQLHVLTTQAGYHW 5393
 Qy 631 -----GPR-----ID-----LLKPD-----641
 Db 5394 ANHFRGRHRAEENAIAGFNMIIVDVVFLVEEDCOLKEETMSEEFEGHGMNDLILIV 5453
 Qy 642 -VSAPGK-----NI-----KSTLN-----654
 Db 5454 GFSAAKGSASLRNIRNQRWVYLNTEACKRPLPFKNFNGVRITDPYQILEYFDQCIHRD 5513
 Qy 655 -----VI-----656
 Db 5514 DVDGIIVDSLTFMMDLETQVVLTAANTQAWGEBFAQFFKILLQEKVVKFAKPVIFTAHV 5573
 Qy 657 -----NKGSTY-----662
 Db 5574 KDEVDAMELKTFPVVKSLKNNGIEAYFSTVVSABRIDLKELEKYSNGMLTEITDEQE 5633
 Qy 663 -GY 664
 Db 5634 LGY 5636

RESULT 11
 ABB58144

ID ABB58144 standard; Protein; 7107 AA.

XX ABB58144;
 XX 26-MAR-2002 (first entry)
 DT XX Drosophila melanogaster polypeptide SEQ ID NO 1224.
 DE XX Drosophila; developmental biology; cell signalling; insecticide;
 KW XX pharmaceutical.
 OS XX Drosophila melanogaster.
 XX WO200171042-A2.
 PN 27-SEP-2001.
 PD XX 23-MAR-2001; 2001WO-US09231.
 PF XX 23-MAR-2000; 2000US-191637P.
 PR XX 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI, 2001-656860/75.
 PI N-PSDB; ABL02247.
 XX
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 DR genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions -
 PT Disclosure; SEQ ID NO 1224; 21pp + Sequence Listing; English.
 XX
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
 CC sequences (AB101840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 7107 AA;
 Query Match 26.6%; Score 1167.8; DB 22; Length 7107;
 Best Local Similarity 7.6%; Pred.No. 7e-07; Mismatches 163; Indels 3642; Gaps 139;
 Matches 324; Conservative 163;
 Qy 1 Y-----PVVL---AD-----TSSSEDALNISDEKVA--ENKEKHENI-----33
 Db 1071 YDVNRGEPAPVITWYQNDKELPEELPSSE-IGNIPYNTKISITVETVKTGIVKIIA 1129
 Qy 34 ---HSAMETS-----QD-----PK 44
 Db 1130 VNEHQDEATVEVNILAPPSPKRGFLDVKVTQKSKLKWKKPEDDGGKPI SAYQVKEFPD 1189
 Qy 45 EKK-----TAVIKEKE-----VSKNP---61
 Db 1190 KQGRWVPLGRTSANDTEFDVKGLOEGHEYOFRVKAINEGESDPLSDSDSIIAKNPDA 1249
 Qy 62 ----VID-----NNTSNEEA 72
 Db 1250 ASKPTPTNIVDNEHMKLWEAPRSDGAPISGVIIIEKKDFSPIDWELLSTNTSVPEA 1309
 Qy 73 KIK---EEN-----SNKSGQ--DYTDS-----FVNK-----93
 Db 1310 TVEGLVEGNIYQFRVAVNK-AGFSDPSDAPTEPHLAKPRNLKPYINRDKMKPIKVRAGOP 1368
 Qy 94 -----NT-----EN 97

```

Db 1369 VKFDVVKGBAPSLTWFLKETELTSTQVRLNIDVNTKLTLDTPDRKQSGQYKLRAN 1428
Qy 98 ----- 97
Db 1429 INGVDAVVEVILIDKPSRPGPIEVSDIHKEGCKLWKRKPDGGLPIGTIVIEKMDTA 1488
Qy 98 -----PKED-----KVY----- 105
Db 1489 TGKAVPAGSVDPKDIKGLDPNHRQYFRVKA NVGSESEPLETSATIAKAPFDVSA 1548
Qy 106 -----YIAEFKDESGE--KAIKELS-----SL 126
Db 1549 PRGLPELEDWDEHNHYKLMKRPPIRDGSGPITNVIIEVMDKSGEFVKA VETSDPYCKGV 1608
Qy 127 KATKVLTY-----DR----- 138
Db 1609 KLEEGQOYKFRVAVNAKAGSPDSEQTNMVAKRPLKPHIDRVNLKPVITGLSISL 1668
Qy 139 -----FNCSALETTP-----DNLDKIKQIEGSSVERAKOY-----PM 172
Db 1669 DINRGEAPKVEWFFNNSVTSDEHSVKIDVINTKF-----FVMAKORSQSGKIITKA 1724
Qy 173 MNHARKEIGVEA-----IDYUK----- 190
Db 1725 TN-----EYGEDEBAELEVTVLGKPKRKPLOVNDITKHSCKLWMEKRPDDGSGPIDYEI 1780
Qy 191 -----SIN-----APF 196
Db 1781 EKLDPHTGOMLPCGKSTEPRAKVIQLHKGAKYKFRVAVNKEGESEDELETERPIAKNPY 1840
Qy 197 -----G-----KNF-----DG----- 202
Db 1841 DEPRPGKPEPTNMDKDFVDLAMPKPDGAPLOKVIQMRKDSGRAMVDSATVPDGC 1900
Qy 203 ----- 202
Db 1901 NGTGVGEHGEYEFRIAVNVKAGSPSDVSKSVIAKRPFLKPHIDRKNLOKKIMRSGQ 1960
Qy 203 -----RGWY----- 207
Db 1961 MLHIDALIKAPPAKVTWYNKTEIKTSDHIKIENEDYKTFIMPKVRA DRGIYVAK 2020
Qy 208 -----SNI-----DTGTD----- 215
Db 2021 NDSGSDIVELEVLCKPSKPGPLAVSNVTAETLHLKMEKPEDGDPIDQYLVERMOT 2080
Qy 216 ----- 215
Db 2081 ETGRKVPVLTKTPEADVTGLTEGKEYLFRVKA VNSGESEPLVTDTITKAKNPFDA DT 2140
Qy 216 -----YR----- 217
Db 2141 PGKPOIVMSGNCDLKWRAPEDDGASITGYIVERKDPNTGKMQALESTPDCARVN 2200
Qy 218 ----- 217
Db 2201 DLINGNKYQFRIMAVNKA GSKSPSDQMTAKDFAPPKIDRTNIKDIITIKAGQHIFD 2260
Qy 218 -----HKAMRI--DD-----DATA-----SMRF----- 233
Db 2261 IKVSGEPATKVMHLNKA RLENDSDSNVINIMESYATKLTVPISKAFHSGKTYTLKAENSG 2320
Qy 234 -----KKEEDL----- 238
Db 2321 RDEASFVJVLDKPPEGPLRVTDVHKEGCKLKNA PLDGLPIDHYIIEKMDVESGR 2380
Qy 239 ----- 238
Db 2381 WLPGRFKESFAELNNLEPSHEYKFRVLA VNTGESEPLTGEQSVIAKNPDEPGKPGTP 2440
Qy 239 -----KGTDKNWL----- 247

```

```

Db 2441 EAVDMDKDHDVLRWPRPIINDGSPITGYVVEKREKGTDK--WIKTEITITPLGEBSKAT 2498
Qy 248 -----SD-----KI----- 251
Db 2499 VPTLNCXEYEFYKALINAGPGSPDASKPIITKPKLAPKIDRKINIRTYNFKSGEPIF 2558
Qy 252 -----P-----HAFNY--YN----- 259
Db 2559 LDINISGEPA PDVTWNNKNSVOTTSFSHIENLEPYNTKYINNPERKDTGLYKISAHNFY 2618
Qy 260 -----GK-----ITVEKYD-D-- 270
Db 2619 GQDQVEFOINITKPKRPGPLEVESEVHKDCKLWKKPKPDGSGEPVESYLVEKFDPDTC 2678
Qy 271 -----GR-----DY-F-----HGM 280
Db 2679 IMLPVGRSDGEYVNDGLVPCHDYKFRVKA NVKEGESEPLETGSIIAKDPFSVPTKGV 2738
Qy 281 H-----IAGIL 286
Db 2739 PEPTDWTANKVELAMPPEPASDGSPIOGYIVEVKKYSPLMKALBTNSPTPTATVOGLI 2798
Qy 287 AGNDTE----- 292
Db 2799 EGNFYQFRVVALNKGSGSPDPSKIFTA KPRYLAKIDRRNLNITLSGTLALDANI 2858
Qy 293 -----QDIKNF-----GLDGI 304
Db 2859 TGEBA PKWELSNYHLOSGKNVTIETPDYTKLVIRPTORSGEYLVATNTSGKDSV 2918
Qy 305 -----APNA--OI-----FSYKMSDAG----- 320
Db 2919 LVNVVITDKSPSPGPIQISDVHKEGCHLWKRKSPDGGPIEFQIDKLEPETGCMTPS 2978
Qy 321 -----SG-----FACDETU--F----- 330
Db 2979 CRSTEPQVDVTGLSPGNEYKFRVSAVNAEGESQVLVGDESI VARNPFDEPGKPENLKATD 3038
Qy 331 -----H-----AIEDSI----- 337
Db 3039 WDKDHDVLA WTPPLIDGSPISCIYIEKODKYGMBALDVPADOCKATIPDLVEGOTYK 3098
Qy 338 ----- 337
Db 3099 FRVSAVNAAGTGESDSTPPIIAKARNKPPIIDRSSLVEVRIKAGOSTFPDCKVSGEPAP 3158
Qy 338 -----KNAVNVSV-----SSG--F----- 350
Db 3159 QTKWLKKKEVYSKDNKVTNVDTNKLK VNSATRSDSGIYTVFAENANGESADAVKVTV 3218
Qy 351 ----- 352
Db 3219 IDKPA PNGPLKVDINSESC TLMNPDDGGOPI DNYYVEKLEDTTGRWIPAGETDGP 3278
Qy 353 ----- 352
Db 3279 VITALKVGSLTFGHKYKFRVAKNRQGTSEPLTTQALIIAKNPFVPTTKPGTPTIKDPFKE 3338
Qy 353 -----TGLVGEKY-----WQA-----I 364
Db 3339 FVDLEWTRPEADGSPITGYVVEKRDKPS PMEKACBESDITNAHVPLDLEGLKYERFV 3398
Qy 365 RALRKAG-----IPNV-----VATGNV-----ATSA--SSSS 389
Db 3399 RAVNVKAGGSPSDATETHVARPKPTPKIDBNFMSDIKIKAGNVFEPDVPATGAPLSKSD 3458
Qy 390 MDLVANNH-----LKMTDPS-----NVTRTAHEDAI 416
Db 3459 W-----THEGNMINTDRVKISNFD DRTKIRILDAKSDIGVYTLTANINGTORHNVKV 3513
Qy 417 AVASA----- 421
Db 3514 TILDAPSVBEPRLANGVSKNSIVLWRP PKDGGSEITHVVEKMDNEAMRWVPVGDCT 3573

```

```
QY 422 -----KNQTVF-----DK----- 430
Db 3574 DTEIRADNLIIENHDYSFRVRAVKNQKOSQPLTTSQITAKDPYSHDPKQGPQATDWGKH 3633
QY 431 ----- 430
Db 3634 FVLEWSTPKRDXGAPISYIIIEKRPKGQWERAANVLGDNKAHVPELTNGGGEYFRVI 3693
QY 431 -VNIGGES-----FKYRNIGAFDKS----- 450
Db 3694 AVNRGSDSDPSSTIICKRFLAPFPDKSLANDITVHAGKRLGWLPIEASPRPLITW 3753
QY 451 ----- 450
Db 3754 LYNGKEIGNSRSGSLFQNELTPEIVSSLSRSGRYTLILKNEHGSFDSAAHATVLDRLP 3813
QY 451 -----KIT----- 453
Db 3814 SPKGPLDITKTRDGGCHLTWNVPDDGGSPILHYIEKMDLSRSTWSDAGMSTHIVHDV 3873
QY 454 -----TNE----- 456
Db 3874 TRLVHRKEYLFRVKAIVNAGESDPLEAVNTIIAKNEFDEPDAPGKPIITDWRDHDLDLQW 3933
QY 457 -----DG-----TKAPS-----KLKFFVIG---K 472
Db 3934 AVPKSDGAPISEYIIQKKEKGSPTWNVHRVPSNKNTTTTIPELTEGOEYFRVIAVNOA 3993
QY 473 GQDQ-----DLI----- 479
Db 3994 GQSEPSFSDMIMAKPRYLPKIITPLNEVRIKCLIFHTDIHFIGEPAPATWTLNSNP 4053
QY 480 -----GLDLRG---KIAVMDR----- 492
Db 4054 LLSNDRSTISIGHHSVVHTVNCQSDSGIVHLLLRNSSGID-EGSFELVLDLPGPPEG 4112
QY 493 -----IYTKDLKNAFKKAMDKGARIMV-----VNTVNY-----YN----- 523
Db 4113 PMEYEEITANSVTISWPKPKONGSGEISSYVIEKRDLTGGGWPAVNVYSKYNHAVP 4172
QY 524 ----- 523
Db 4173 RLLEGTMYELRVMAENLQGRSDPLTSQPVVAKSQYTVPGAPGKPELTDSDKNHITIKWK 4232
QY 524 -----RD-N---WTEL-----PAMGYEAD-----EG 540
Db 4233 QPISNGGSPITGYDIERRDVTNGRWIKINGQPVPTAEYQDDRVTNSHQYQVRISAVNAAG 4292
QY 541 T-----KSQVPS-----ISG----- 550
Db 4293 NGKTSEPSAIFNARPLRKPRFYFDGLIGKRIKVRAGEPVNLNIPISGAPTPTIEWKRGD 4352
QY 551 -----DP----- 552
Db 4353 LKLEGRKISYETNSERTLFRIDDSNRDRSGYTVTAANEFKGTADIEVIVWDKPSPE 4412
QY 553 -----GVKLW----- 557
Db 4413 GPLSYTETAPDHISLHWYSPKDDGSDITGYIIETFERGVDDKVPVPGTCPTNPTVKNL 4472
QY 558 ----- 557
Db 4473 VEGKYVFRIRAENIYGASEALEGKPVLAKEFPDPPGAPQSPTISAYTFNSANLEWHPPD 4532
QY 558 -----NM-----IN----- 561
Db 4533 DCGKPTGYIVERRRERGGEWIKNNYPTPTNTSYTVSNLRDARGARYFRVLAINEAGPQHP 4592
QY 562 -----PD-----KKEVKN-----NKE 574
Db 4593 SKPSDPMTAEHQRYRPDPPEPKPDRIITRNGVTLWSRPPRTDGSRIKGYVEMRPKNKG 4652
```

```
QY 575 DFK---D-----KLEOY-----YPIDME----- 589
Db 4653 DWKTVDIPINSTVYTVPSLSKEGEYSFRVVAENEVGRSDPSKPSQPTIIEQPNKPCME 4712
QY 590 -----SF-----NS 593
Db 4713 LGKVRDIVCRAGDDFSIHVPYLAPPKPNAFWYSNDNMLDDNNRVHKHLTDDAASVVVKNS 4772
QY 594 -----NKNVGDG- 601
Db 4773 KRADSGOYRLQLKNTSGFDTATINVRVLDREPSPTRLRADEFSGDSLTYWNPEN--DDG 4830
QY 602 -----KEIDF----- 606
Db 4831 GSAIQNYIIIEKEARSSTWSKVSFCTVPFVRIRNLVLNKEYDFRVAENKYGQSDPANT 4890
QY 607 -----KF-APDT-----DK-- 614
Db 4891 SEPTLARHPDIPNTPGIPHGIDSTEDSITIAWTKPKHDGSPITGYIIEKRLSLDDKWT 4950
QY 615 -----EL-----Y--KEDIIV----- 623
Db 4951 KAVHALCPDLCKIPNLIIENAEYEFRAVNAAGQSAYSGSSDLIFCRRPPHAPKITSDL 5010
QY 624 -----PAGSTSW-----GPRI----- 634
Db 5011 SIRDMTVIAGDEFRIPTVYPYHASPRPTASWSLNGLEVIPGERIKFDSNDYASMYNKSAR 5070
QY 635 -----DLLKP-----DVS----- 643
Db 5071 DETGSYITLTNNKSGSDTASCHVTVDRLPPQGPLNAYDITPDTCTLAWKTPLDDGGSP 5130
QY 644 ----- 643
Db 5131 ITNYVVEKLDNSGWKISSFVRNTHYDVMGLEPHYKYKYNFRAENQYGLSDPLDIIBPI 5190
QY 644 -----APGK----- 647
Db 5191 VAKHQFTVPDEPGPKVIDWDSGNVTLIWTRPLSDGSGRIQGYQIEYRDILNDSSWNAVD 5250
QY 648 -NIKST---LNIVINGKSTY----- 662
Db 5251 YIIKDTKQVLYNLING-SEYEFRIKAKNAAGLSKSPSSSLRFLKLGKFTVPSPPGAPQVT 5309
QY 663 -----GY 664
Db 5310 RVGKNYVDLKWKEPLRDGSGRITGY 5334

RESULT 12
ABB98574
ID ABB98574 standard; protein; 2835 AA.
XX
AC ABB98574;
XX
XX 14-JAN-2003 (first entry)
DE Dextran saccharase, DSRE.
XX
KW Dextran saccharase; enzyme; cytostatic; dermatological; antiseborrheic;
KW DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;
KW microflora regulation; intestinal transit; mineral assimilation;
KW colon cancer; acne; dandruff; body odour.
XX
OS Leuconostoc mesenteroides NRRL B-1299.
XX
FH Key Location/Qualifiers
FT Peptide 1..40
FT Protein /label= Signal_peptide 41..2835
FT Peptide /label= Mature_protein 423..439
FT /label= SEQ_ID_6
```

FT Peptide /note= "This sequence is specifically claimed in Claim 7"
FT 478..501
FT /label= SEQ_ID 7
FT /note= "This sequence is specifically claimed in Claim 7"
FT 519..539
FT /label= SEQ_ID 8
FT /note= "This sequence is specifically claimed in Claim 7"
FT 560..571
FT /label= SEQ_ID 9
FT /note= "This sequence is specifically claimed in Claim 7"
FT 631..645
FT /label= SEQ_ID 10
FT /note= "This sequence is specifically claimed in Claim 7"
FT 1014..1021
FT /label= SEQ_ID 11
FT /note= "This sequence is specifically claimed in Claim 7"
FT 2120..2138
FT /label= SEQ_ID 12
FT /note= "This sequence is specifically claimed in Claim 7"
FT 2161..2184
FT /label= SEQ_ID 13
FT /note= "This sequence is specifically claimed in Claim 7"
FT 2202..2214
FT /label= SEQ_ID 14
FT /note= "This sequence is specifically claimed in Claim 7"
FT 2243..2250
FT /label= SEQ_ID 15
FT /note= "This sequence is specifically claimed in Claim 7"
FT 2315..2322
FT /label= SEQ_ID 16
FT /note= "This sequence is specifically claimed in Claim 7"
FT 2689..2696
FT /label= SEQ_ID 17
FT /note= "This sequence is specifically claimed in Claim 7"
FT 1981..1142
FT /note= "Catalytic domain, SEQ ID 1. This sequence is specifically claimed in Claim 4"
XX FR2822163-A1.
XX 20-SEP-2002.
XX 19-DEC-2001; 2001FR-0016495.
XX 16-MAR-2001; 2001FR-0003631.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Bozonnet SAM, Remaud SMMC, Willemot RML, Monsan PEF;
XX WPI; 2002-715213/78.
XX N-PSDB; ABQ80961, ABQ80962.
XX New glycosyl transferase enzymes, containing glucan bonding and
XX catalytic domains and producing alpha-(1-2) branched dextrans, useful
XX in probiotic, pharmaceutical or cosmetic compositions
XX Claim 6; Page 65-74; 82pp; French.
XX The present sequence is a novel dextran saccharase, DSR-E, from
XX leuconostoc mesenteroides NRRL B-1299. The saccharase has glycosyl
XX transferase activity suitable for producing dextrans having alpha(1-2)
XX branches from sucrose, p-nitrophenyl-alpha-D-glucopyranoside,
XX alpha-D-glucogalactose, D-nitrophenyl-alpha-D-glucopyranoside or
XX alpha-D-galactopyranosyl-sucrose. The dextran saccharase is useful in
XX probiotic, pharmaceutical or cosmetic compositions. The dextrans and
XX related compounds having alpha(1-2) bonds, produced using DSR-E, may be
XX involved in signalling/cellular recognition processes in vivo
XX (specifically in regulation of microflora in the intestines or on the
XX skin); and are potentially useful for improving intestinal transit,
XX increasing assimilation of minerals (e.g. calcium and/or magnesium),
XX preventing cancer of the colon and combating skin problems such as acne,
XX dandruff and body odour.

XX SQ Sequence 2835 AA;
Query Match 26.5%; Score 1165.4; DB 23; Length 2835;
Best Local Similarity 10.8%; Pred. No. 2.2e-08;
Matches 285; Conservative 160; Mismatches 202; Indels 1986; Gaps 120;
QY 2 PVV--LADTSSSEDA--LNI-----SPKEVAENKEGK-----30
DB 196 PVANKADTSSIHQPLDPTNPTDKSANLVSTQKSIDNQVKSTETSHLOENKITYFL 255
QY 31 -----ENHSAMB-----38
DB 256 DDNGVKKNPFAIIDGKLVYFDKTSGETLANAPQVTKGLVINDAHNAADHLTDANFTNV 315
QY 39 -----TSQDF-----KEKTVAV-----50
DB 316 DGYLTANSWVRPKDILKNGTWTPTTAEDFRLPLMSWMPDKNTQVAYLYQVQSGMLPDD 375
QY 51 -----IKEKVS-----KRPVID-----64
DB 376 VKVSDNDNMSTLTDAAMTVQKNIESRIGVSGKTMWKODMKNKLIDSQAMNIDSESKND 435
QY 65 -----NNT-SNEAKIKE-----ENSNK--81
DB 436 HLGQALLVNDKTPNANSDYRLNRTPTQTGQITDPSKQGYEMULANDVNSNPV 495
QY 82 -----SQGYTDSF-----V91
DB 496 QAEQLMLHYMNIGITIAQNDPTANFDGYRVADVANDVADLQIAGDYFKAAYGTGKTEA 555
QY 92 NKN-----94
DB 556 NANHHISILEDWNNDSAYIRAHGNNQITMDPRHLLAKTALNPLAAGLEPLINTSL 615
QY 95 -----TEN--PKK-----EDKV-----VYI107
DB 616 VKRGKDATENEAQNVAFIRAHDESEVQVIAQIHKDKINTKSDGLTTPDEIKQAFITYN 675
QY 108 A-EFK-DKESGEKAKE--LSSLKQT--KVLV-----TYD-----136
DB 676 ADELKADKEVTAAYNIPASVAVLLTNKDTVPVRYVGDLPSSDDGYMSQKSPYDAITSLK 735
QY 137 -RI-----FNGSAIETP-----DNL-----151
DB 736 SRIKYVAGGQSMNTYLHECFDPAKNETKPGQGLTVRYGKAMTADBLGNSDTRQOIG 795
QY 152 -----151
DB 796 LVINNKPFLLNDDBOIVLWNGAAHKQOAYRPLMLTTKSGIQTIDKXGAPVYTTNDAG 855
QY 152 ---DKI---KQIEGI-----SSVERAQKVQ-----170
DB 856 LIFKSDMVYGVSNPQSGYFAAWVPVGASDQDARTOSSQGETKGDVYHSAALDSNVI 915
QY 171 -----PMN-----174
DB 916 YEGFSNFQAMPEKNDFTFNKIAQNAKLFKDGLITSELABQVRSSTDSNFLSVIONGY 975
QY 175 -----HARKEIGVEAIDYK-----190
DB 976 AFTDRIVGVNTPPTKYGTVQDQLDLSRALHQA---GIQAINDWPPDQIYNLPGEOIVTAV 1032
QY 191 -----SINAP-195
DB 1033 RTNCGKYVDYSVINNTLYDSRTVGGEGYQKFGGLFLDQKKDYPSLFETKQISTNQPM 1092
QY 196 -----F-GKNPQDRGM--VYS-----NI--DNC-----213
DB 1093 NPDKIKEMSAKYNFNGSNIGRGAMVYLKDMATNQVFNVSNDNPLPKQLIGEXTSGCFI 1152
QY 214 -----TDYRHRKAMRID-----DAKASMFKEKEDLGTGKNTYVLSKIP---252

Db 1153 TENGKTSFYSTSGYQAKDTFIODGTNNWYFDNAGYMLTGKQNI--HDKNYTF--LPNGV 1207

Qy 253 ---HARN---YYN---GGKITVEKYDD---GR---D--- 273

Db 1208 ELQDAYLFDGNOEFYNNKAGEQVMQYQDSQONWHYFFENGMAIGLTEVPNADGTHVT 1267

Qy 274 -YFDPHGMHIAG-----ILAGN-- 289

Db 1268 QYFDANGVOIKGTAIKQNNQLRYFDEATGNMNVNSWGOLADKSWLYLNAQGAUVTGNQK 1327

Qy 290 -DTEQ-----DIK----- 296

Db 1328 IDGEYFNADGKQVKGNAIDNNGDQRYDGDGKGMVNVNSWGEPLDGSWLYLNDKGIATV 1387

Qy 297 ---NFGIDGAPNAQIF-SYKMSYDAGSGFAGDE-----TWP- 330

Db 1388 TGRQVINQVNFPGNDG---KQIDAPKLLSD-GSWVYLDKGLITTGAKVINGLNMF 1442

Qy 331 ---HAIE-----DSI--KHNV-----VVSUSSG-----FTG----- 352

Db 1443 DKQHOIKGDASTDANGKRHYDYDKNDGHLVTNSWGEPLDGSWLYLEEGDAVGTORVIDG 1502

Qy 353 ---TGL--VGEK--Y----- 360

Db 1503 KTRYFDEDEKQIKNSLKTLANGDKIYLDGDGVAATGLQHVGDKIMYFDEDEGKQVVGKFS 1562

Qy 361 ---W-----QAIRALKRAGIPMVVATGNATYSAS----- 386

Db 1563 AKDGSWYLYNQDGAAGVSSINGQSLYFDQDGKQVKNVNRNSDGTNNYTGLTGKLT 1622

Qy 387 ---SSW---DL-----VAN-----NHLKM--TDT 403

Db 1623 QDFGELPDGSWYLYDAQGHTVTGAQIINGQNLVFKADGQVKGHAYTDQLGHRFYDPDS 1682

Qy 404 GN-----VTRTAHE-----DAIVASAKNQ-TVE-----F 428

Db 1683 GMLNRPFOITPGVWAFGADGVAITGQHDINGQKLPFDETYGVQVKGSOITIDGTLSP 1742

Qy 429 DK----- 430

Db 1743 DSOTGNQKRVOTLLPOAGHYITKNGNDWQYDTNGELAKLRQDSNGKLRYPDLTTGIA 1802

Qy 431 ---VNIGES--FK-----YR----- 441

Db 1803 KGQFTVIGQETYYFSKDHGAQLPMTVEGHYGTITLKGQDQTKTAWVYRDQNNITLKL 1862

Qy 442 -NI-G-AFFDKSKITTNE--DGTKAPSKLKFVYI--GKG-----QDQDLIGL 483

Db 1863 QNINGTLOFFDP---YTGEQLKGGVAKYDDKLFYFESGKGNLSTVAGDYQDGHYISQDG 1919

Qy 484 R-----GKIAVMD-----RIYTKDLK----- 499

Db 1920 QTRYADKQNLVKGLVTNGALQYFDNATGNQIKNQQVTDGKTYFYDDKNGEYLFNT 1979

Qy 500 ---NAP----- 514

Db 1980 LDMSTNAFTKNVAFNHDSSFDHTVDGFLTADTWYRPSILANGTTWRDSTDKMRPLI 2039

Qy 515 VV-----NT-VNYN--RDNWELPAMGY--BADE-----GTSQVFSIGDD 552

Db 2040 TVWPNKVVQVNYLNFMKANGLLTTAAQYTLHSDQYDLNQAAQDVQVVAIERIASEHGT 2099

Qy 553 G-----VK---LWNM-----INP----- 562

Db 2100 WLQKLLFESONNPNPSFVKQOFIWNKDSYHGGDAWFGQGYLKGNNPLTPTTNSDYRQ 2159

Qy 563 ---DKTEV-----KENNKEDF--KDK 579

Db 2160 GNAFDPLANDVNSNPVQVAENLNLWLYMNFGITAGQDDANFDSIRIDAVFIHNDT 2219

Qy 580 LEQYV----- 584

Db 2220 IQRTYDLRDYVQVQOSEAKANQHISLVEAGLDAGTSTIHNDALIESNLREAATLSLTNE 2279

Qy 585 ---P-----ID----- 587

Db 2280 PGKNKPLTNMLQDVGDTLTIDHTQNSTENQATPNYSIIHHAHDKGVQEKVGAATDTATGA 2339

Qy 588 ---MESF-----NS-NKENV-----GDE----- 601

Db 2340 DWTNFTDEQLKAGLELFYKQORATNKYNSYNIISIVALMLTNKDTVPRMYGDMYQDDG 2399

Qy 602 ---KBID-----FKF----- 608

Db 2400 QYMANKSIYYDALVSLMTARKSYVSGQTMDSVDNHGLLKSVRFGKDAMTANDLGTSA 2459

Qy 609 ---AP-----DTDK-----ELYKEDI----- 622

Db 2460 EGLGVIIIGNDPKQLQNDSDKVTLDGMAAHKNQKYRAVILITTRDGLATFNSDQAPTAM 2519

Qy 623 ---VPAGST----- 628

Db 2520 QGTLTFSNOEINGQDNTQIRGVANPOVSGYLAHVVPVSGASDNQDQARTATTENHDGKVL 2579

Qy 629 ---SWG-----P----- 632

Db 2580 HSNALDSNLIYEGFSNFPQKATTHDELTVNVIKADVFNWNGITSPFEMAPQYRSSGDH 2639

Qy 633 ---RIDLLK-P-----DV-----SAP 645

Db 2640 TFLDSTIDNGVAFTRDYDLGFTPTKYGTGDLRATTIQLHANNQVNVADVVDNQVYNLP 2699

Qy 646 GKNIKSTL----- 653

Db 2700 QKEVVSATRAGVYGNDDATGFGTQLYVTNSVGGQYQEKYAGQYLEALKAKYPDLFEGKA 2759

Qy 654 ---NVIN-GKSTY-----GY 664

Db 2760 YDYVKNVANDGSPYVTLTSHGDRSIPADVAIKQWSAKYMNVTNLGNGMGY 2812

RESULT 13

ABG74786

ID ABG74786 standard; Protein; 31267 AA.

XX

AC ABG74786;

XX

DT 05-JUN-2003 (first entry)

XX

DE Human RGS11 protein.

XX

KW RGS11; human; screening; cardiant; antianginal; gene therapy;

KW heart disorder; cardiac ischaemia; heart failure; angina.

XX

OS Homo sapiens.

XX

PN WO2002103355-A1.

XX

PD 27-DEC-2002.

XX

PF 17-JUN-2002; 2002WO-JP06019.

XX

PR 18-JUN-2001; 2001JP-0183038.

XX

PA (TAKE) TAKEDA CHEM IND LTD.

XX

PI Koyama N, Tanida S, Yamamoto K;

XX

DR WPI; 2003-167557/16.

DR N-PSDB; ABX13540.

XX

PT Screening compounds regulating RGS11 expression and activity for

PT prevention and treatment of heart disease

XX

PS Claim 1; Page 59-261; 321pp; Japanese.

XX

CC This invention describes a novel method for screening compounds for their
 CC ability to regulate the activity and expression of human RGS11 and its
 CC partial peptides and salts, by observing the expression or activity of
 CC RGS11 in the presence or absence of the test compound. The products of
 CC the invention have cardiant and antianginal activity and can be used for
 CC gene therapy. The methods and compositions are useful in the prevention,
 CC treatment and diagnosis of heart disorders such as cardiac ischemia,
 CC heart failure and angina. This sequence represents the human RGS11
 CC protein described in the disclosure of the invention.

XX
 Sequence 31267 AA;

Query Match 26.5%; Score 1165.2; DB 24; Length 31267;

Best Local Similarity 8.8%; Pred. No. 0.00025;

Matches 312; Conservative 167; Mismatches 170; Indels 2908; Gaps 139;

```

QY 1 YPVV-----LADT-----SSED-----ALNT--SDKEV----- 23
DB 9448 LPVEDCAISGSEPIVSWYKDKPLKSPNVQTSFLDNTATLNIFKTRSLAGQVSGTA 9507
QY 24 -----AENK-----EKH----- 30
DB 9508 TNPISASSASARLLITEGKNPPFDRIAPVDAVNGESADFECHVTGTQPIVSWAKSR 9567
QY 31 -----EN-TH-----SAMETSOD-----F 43
DB 9568 EIRSGGKQIYSLYENSAHLTVLKVKGDSQGYTCVAVNEVGSDCTAQLNIKERLIPSF 9627
QY 44 KEKTAIVIKKE-----VSKNPVT---DNN-----TSNEAKIK----- 75
DB 9628 TKRLSETEETGNSFKLEGKRVAGSQPIVAMVKNNIEIQTSPNCEITFKNNTLVLOVRK 9687
QY 76 -----EENSNS----- 83
DB 9688 AGANDAGLYTCVNSDAGSALCTSSIVIKPKKPPVPOHLLTPTVSSEGVQLSCHVQG 9747
QY 84 -----GDY-----TDS--- 89
DB 9748 SEPRIQWLKAGREIKPDRCSFSAAGTAVLELADVAKADSGDYVCKASNVAGSDTTKS 9807
QY 90 -----FVN--- 92
DB 9808 KVTIAKEPIQFTKRIQNIIVSEHQATECEVSPDDAIVTWYKGPTLETESQKVFRRDNG 9867
QY 93 ----- 92
DB 9868 RCHWYTHNVTPDDEGVYSVIALRPEGEARSTAEYLTTKEIKLEKRPDIIDSRVPI 9927
QY 93 ----- 92
DB 9928 TMPIRAVPEEIPVVAAPRILLLPTPEKPKPPKRIESHESHERKVPKAVPEKKAPPPKV 9987
QY 93 -----KNT--- 95
DB 9988 IKKPVIEKIEKTSRMEEEKVQTVKPEIKPAIPAPPEKPKPEAGVPKPTPSIEABR 10047
QY 96 -----ENPKEDKVVY-----IAFKD---KES---GEKAIKE--LSLKTAK 130
DB 10048 KRLRPGSGEKEPPDPAFTYQKAVPLKFKVKEIKDIILTESEFVSSAIFECTVSPSTA 10107
QY 131 VLYTYDRIFGNSAIEETP-----DNLDK----- 153
DB 10108 TTMMD-----GSNIESPKGRFIADGKDKRKHIIIVQLSDAGEYTCVLRNGKEKTSIAK 10163
QY 154 -----IKQIE-----GT--- 160
DB 10164 LVVEELPVRFVYKTLREEVTVVKGQPLYLSCELNKRDEVVVRKDKIIVKEKGRIVPGVIG 10223
QY 161 -----SSVE-----RAQYQVM--- 172
DB 10224 LMRALTINDADDTAGTYTAVTVVENANNLECGSCVAVVEIVDMLVKPIRDQHVKEKGTAI 10283
QY 173 -----NNH----- 175

```

```

DB 10284 FACDIADKTPNIKWFPGYDEIAPBPNDKTEILRDGNHLYLKIQAHPEDIAEVAVEIEGK 10343
QY 176 -----AKK----- 178
DB 10344 RYPAKLTJGEREVELLKPEDVTIYKESASPDASEADI PGQWKLKGLLRSPPTCEI 10403
QY 179 -----EIG-----VEE-AIDY---LKSINAP----- 195
DB 10404 KAEGGKRLTLHKVKLDQAGEVLYQALNATITALLTVAEIEDVDPVLPKDVVPERROAR 10463
QY 196 F-----GKN-----FD-----GRGVISNIDTGD----- 215
DB 10464 FECVLTREANVYWSKSPDIKSSDKFOIADGKHILVINSQFDDDEGVTAAYEGCKTS 10523
QY 216 -----YHKAMRT----- 223
DB 10524 ARLFVTGIRLKFMSPLBEDQVKEGTATFVCELSHEKKNVVMFKNDAKLHSTRVLISSE 10583
QY 224 -----DD----- 225
DB 10584 GKTHLEKKEVTLDDISOIKAQVKELSTAOQKVLEADPYFTVKLHDKTAVEKDEITLKC 10643
QY 226 -----DAKASMRP-----KEDDLK-----GTDKN--- 244
DB 10644 EVSKDVPV-KWFKDGEIIVSPKYSIKADGRLRIKLIKAKDLKCKGGEVVCDCGTDKTKAN 10702
QY 245 -----Y-----W-----LSDKI 251
DB 10703 VTVEARLIKVEKPLYGVEVFGETAHFEIELSEBDVHQWMLKQGPLTASPDCEIIEGK 10762
QY 252 PHAFNYVN---G-GKTIYE-----KYDD----- 270
DB 10763 KHLILHNCQLGMEVSVFOAANAKSAAHLVKELPLIFITPLSDVVKVEKDEAFCEV 10822
QY 271 -----GRDYF---D---PHGM-----HIAG--I 285
DB 10823 SREPTFRMLKGTQEIITGDDFFELIKDCTKISWYIKSAAFDEAKYMEAFADKHTSGKI 10882
QY 286 LAG-----ND-----TEDD----- 294
DB 10883 IEGIRLKEPLTKDVTAKESAVFVELSHDNIRVKVKFKDQRLHTRSVMDDEGKTH 10942
QY 295 -----IKANGT----- 301
DB 10943 SITRKDISIDTQIRVEAMGMSSEAKLTVLEGDPYFTGKLQDYTVGEKDEVIILOCEISK 11002
QY 302 -----DG--IAP--NA-----QIPSYKM-----YS-DAGSGFAGDETM-- 329
DB 11003 ADAPVKFKDQKEIKFPSKNAVIAKADGKKRMILKALKKSDIGQYTTCDGCT---DKTSGK 11058
QY 330 -----FPAI-----EDSI----- 337
DB 11059 LDIEDREIKLVPRPHSVVEVMEETARFETIESEDDIHANWKLKGBALLQTPDCEIKEEGK 11118
QY 338 -----KKN-----VDV-----VSVSG----- 349
DB 11119 IHSVLHNCRLDQGTGVFOFQANYSAAHLVKRPRVIGLARPLKDYVTVJAGETAFFDEL 11178
QY 350 ----- 349
DB 11179 SYEDI PVEMWYLGKKLERSDKVPRSEKVTTLTRLDVLYLEDADEVQLTAKDPKTHANLF 11238
QY 350 -----FT-----GT-GLVGEKY----- 360
DB 11239 VKPEPVEFTKPLEDQTVBEGATAVLEGEVSRENAKVMFKKGTETLSKKYVEIVADGRVR 11298
QY 361 ----- 360
DB 11299 KLVIHDCPTEDIKTYTCDADKFTSCNLNVPPHVEFLRPLJDLQVREKEKARECELSR 11358
QY 361 -----W-----QAIRAL-----RKAGIPNV- 375

```

Db 11359 ENAKVKFKDGAIEIKKKYDIISKGAVRILVINKLLDDEAEYSCVETARTSGLTVL 11418
 Qy 376 ---VATGNYATSASSS---W---D----- 391
 Db 11419 EEEAVFTKNLANIEVSETDTIKLVCEVSKPGAIEVIWYKGBEIIETGRYEILTEGRKRL 11478
 Qy 392 LVANNHLKMTDTGNV---TRT---AAHE----- 413
 Db 11479 VIGNAHLE--DAGNYNCLRPSRTDGGKVKVHELAAAEFISKPONLEILGEKAEFVCSISK 11536
 Qy 414 ---DAIA-----VASA----- 421
 Db 11537 ESFPVQWKDDTLESBGKDIADGKKRVLVVKDATLQDMGTYYVMVGAABAAHATVI 11596
 Qy 422 ---KNQVFPD-KVNIAGESFKY-RNIGAPFDKSK----- 451
 Db 11597 EKLRIWPLKTRVKEQGVFVNCVNTGAKAKWFRNEAIFDSSKVIILQKDLVYTLR 11656
 Qy 452 ---ITTND----- 457
 Db 11657 IRDAHLDDQANNVNLNHRGENVKSAAANLIVEEBDLRIVEPLKDIEIWEKKSVTFWCKV 11716
 Qy 458 ---GKAPSKLKFEV-----IKGQD----- 475
 Db 11717 NRLNVLTKWTKNGEVPDNRSYRVDKYKMLTIKDCGFPDEGEVIVTAGQDKSVAELL 11776
 Qy 476 --- 475
 Db 11777 IIEAPTEFVHLEDQTVTFDDAVFSCLSREKANVKWYRNGREIKGKKYKFEKDGSIH 11836
 Qy 476 ---ODLI---GLD----- 482
 Db 11837 RLIIKDCRLDDECEYACGVEDRKSRARLFEVIEIRPPQDILEAPGAOVVFLAELNK 11896
 Qy 483 ---LR-----GKI---AVMD-----RIYTKD----- 497
 Db 11897 DKVEVQWLENNVVVQGDQKQWSEKIHRLQICDKPRDQGEYRFAIKDKARAKLELA 11956
 Qy 498 ---LKNA-----FK----- 503
 Db 11957 AAPKIKTADODLVVDVGKPLTMVVPYDAYPKAAEWFKEKEPLSTKTIDTTAEQTSFRIL 12016
 Qy 504 ---KANDKG----- 509
 Db 12017 EAKKGDGKRYKIVLQNKHGKASGFNLKVIDVPGVRNLEVETETFDGEVSLAWEEPDTG 12076
 Qy 510 --- 509
 Db 12077 GSKIIYVVERRDIKRTWVLTADRAESCEFTVTGLQGGVEYLFVRSARNRVGTGEPVE 12136
 Qy 510 ---ARA---TMVN-----TVNY----- 521
 Db 12137 TDNPEARSKYDVPGLPLNVITDVRFGVSLTWPEPYDGGAEITNYVIELDKTSIRW 12196
 Qy 522 ---Y---NR----- 524
 Db 12197 DTAMTVRAEDLSATVTDVVEGGEYSFRVRAQNRIGVGKPSAATPFVKVADPIERPSPVNV 12256
 Qy 525 ---DNW----- 527
 Db 12257 LTSSDOTQSSVQLKWEPLKDGSGPILGYIIBERCEBGKNWIRCNMKNLVPETYKVTGLE 12316
 Qy 528 ---TEL-----P 531
 Db 12317 KGNKYLRYSAENKAGVSDPSEILGPLTADDAFVEPTMDSAFKGLGVIVPNPITILVP 12376
 Qy 532 ANGYE-----ADEG-----TKSQVFSIS 549
 Db 12377 STGYPRPTATWCGDKVLETGDRVKMKTLISAYABLVISPSERSDKGIYTLKLENRVKTLIS 12436
 Qy 550 G-----DDG-----VKLWNMI 560
 Db 12437 GEIDNVNVIARPSAPKELKFGDITKDSVHLTWEPDPPDDGSGPLTGYVYVVEKREVSRTKTVK 12496

Qy 561 ---N-----PDKKTEVK-- 569
 Db 12497 MDFVTDLFTVVDLVQCKEYLFVKARNKCGPGEPAYVDEPVNMSTPATVDPDENVKWR 12556
 Qy 570 -RNNKEDE-----RDK-----LEQ--YY 584
 Db 12557 DRTANSIFLTWDPKPNKDGSRIGYIVVERCPRGSDKKWVACGEPVAETKMEVTGLEEGKWY 12616
 Qy 585 -----P-----IDM-----ESP----- 591
 Db 12617 AYRVKALNRQGASKPSRPTBIEIOAVDQEAPEIFLDVKLLAGLTVKAGTKIELPATVTGK 12676
 Qy 592 -----NSNKP----- 596
 Db 12677 PEPKITWTKADMILKQDKRITTIENVPKSTVTIIVDSKRSDTGTVIIIEAVNVCGRATAVVE 12736
 Qy 597 -NVGDE--KEIDFKFA-----P-D-----TDKELY----- 617
 Db 12737 VNVLDKPGPPAAFDITDVNTESCLLTWNPPRDDGSGKITNVVVERRATDSEVWHKLSSTV 12796
 Qy 618 -----KEDI-----IV-----PAGSTS----- 629
 Db 12797 KDTNFKATKLIPNKEYIFRVAENMYGVGEVQASPIAKYQFDPGPGPTRLPSPDITKD 12856
 Qy 630 -----WGPRIDLLKPD----- 641
 Db 12857 AVTLTWCEPDDGSGSPITGYWVERLD---PDTDKWVRCKMPVKDITYRVKGLTNKKKY 12912
 Qy 642 -----VSAPGKNIKST-----LNVIN-----GKS 660
 Db 12913 RFRVLAENLAGPGKPSKSTELIKPIDPPSPRWLEVINITKNTADLKWTVPKDGGS 12972
 Qy 661 ---TY-----GY 664
 Db 12973 PITNYIVEKRDVRRKGW 12989
 RESULT 14
 ABB66417
 ID ABB66417 standard; Protein; 4345 AA.
 XX
 AC ABB66417;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 26043.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL10520.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX

PS Disclosure; SEQ ID NO 26043; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins

CC (ABBS7737-ABBS7072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

CC

XX Sequence 4345 AA;

Query Match 26.5%; Score 1164.4; DB 22; Length 4345;

Best Local Similarity 9.7%; Pred. No. 1.2e-07;

Matches 302; Conservative 159; Mismatches 184; Indels 2456; Gaps 131;

1 YPVVL-----ADTS-----SEDAL--NISD-KE 21

97 YTVAVRTEPAGSGDSSDLLKSLDLEIFPKPCGYLRINDAKLYDTLDELNDSDSE 156

22 KYAE-----NKE-----KAE----- 31

157 KTKEDYVDNLANESESONYDNMHPKSSDPVNDLTKNLRFAFHGDLISEVCPOQETP 216

32 --NI----- 33

217 WVLNKKGILSAFQNTMRFPVDANTETDVSGCQVQVLEDDTSSVYTRKTDINSC 276

34 -----HAMERSQ--DFKEKTA--VIKEKVSKNPIYDNNTSNEAKIE-----E 77

277 KQRYATHSVLQTPYTFRDKTIPILKSQ--HCNLTIDNNVKE--IKCLETHLLVP 331

78 NSNKSQGYTDS-----FVNKTE-----NP--KKEKV-- 104

332 FSNASSGALLTSTSLKLDGVESYAGEFLQNPFLVERRATLVDPHPRAYKPSDELKA 391

105 --VYIA-----EFKD----- 112

392 ARELLVENCVRGPNIQREFIDVFTNPLOTSKSLDYKTLVLLQBSASTCEGRNHLLES 451

113 ----- 114

452 LPFIQSTASYKVMQTLTFHSLIEYAKNRDLAEYTLGATAVNHSECKHAECEENLRVQ 511

115 -----SGEKAIKE-----LSLSLKTKV----- 131

512 IINLLETFLNLNLFKGERRTREKRVILKGLGNIGVSSAFABQLOMIIREDAPVDI 571

132 -----LTYT-----DRIFNG--SAIE-- 145

572 RLHGILAFRRVDCARHRSYFLDNYGNYTLNSELRYSLQAMRCDIYISVGTIKLLEHE 631

146 -----TTP----- 148

632 EINQGSFVWSHLTLAKSNSPVRIEAQGLINDELSEFKMDIRKFSRNYEHSILFPEY 691

149 ----- 148

692 NFGTTTANVIFGTDSYLPRIASVNFADLFGQSVNFEFTTARAGELEBLANAFGPXGP 751

149 -----DNL-----DK 153

752 LSGQLRKRLKSLFLNRMWLGESAEEDDTLENLSDNLRLKRAQOQASAESEVEDVDYF 811

154 IKQIGIS-----SVERAQKQVPM-----NHAKKEIG-----VE- 183

812 EESLEGNRKQKRDVSTTRKQELIDRVNDSLGLKYDYNNPRAQFGLRVFGNDLRYFNES 871

184 --EALDYLSINAPF-----GKNF----- 200

DB 872 LVEWALAAKFN--PFOQAKVLSGKEFTYTKSRVFLDASTVPLAVGLPLAIHAFGASSI 930

QY 201 DGRGMVLSNT---DTGTDYR-----KARID----- 224

DB 931 DLR--VSGNIDEMDPPTDW--HFVDEGQFKPSVSDVITTTQTDPMFWEOSIKVKSNDYSN 987

QY 225 ----- 224

DB 988 SELVAKLKVRGRNLVSPFSFPLPRDKNEIFSVRSSELLVQKEEQLPOAGIANRSANSTCTW 1047

QY 225 -----DDAKA-----SNRFKXEDYKTDKNY-----W 246

DB 1048 PVLDOAIGLQWCHSYVSDLSNATEIYPSLLAGPLNFSLLKXSDL--SAKKVVFEXKW 1105

QY 247 LSD-----KIPAHN-----YNG-GKITV-EKYD--- 269

DB 1106 --DOOEEDNNFSLVPTTSPGSKVPRVLVAVYKVDAFNASVAFVNGPRVSAGCSYDGNP 1163

QY 270 DGR--D-YFDPHG-----MH-----IAGI-----LAG-----N 289

DB 1164 DFRRLDIYLDITNGNRSJDLGHELRHODFTAMITNPRMLAINGVNITGLGVKVNEN 1223

QY 290 DTEQ-DIK-----NF-----NG- 300

DB 1224 GIKQHDVLSFETKKLQAVIKGNVQSEITSTNMITIKYRQANKIEINPAGKLVNNGD 1283

QY 301 -----DGIAPN-----AOIF 311

DB 1284 KSKTEYRGNMKLQTSAYPKLNFASESTWLSLQGTGMITVNNAPDVNPNRYTSLVRLIF 1343

QY 312 S-----YK-----MYS----- 317

DB 1344 ARSHSEDSFLDGTOTRASLELKLPRSKI DYRLVYKHEKHNGTEHNIVGLTKPEKEI 1403

QY 318 -----DA-----GSGFAG----- 325

DB 1404 TGLFSVHLPRNRLFAIDAYNNVTVPFRNSCTASLKNVEKATKOYIIFINSWTFGHSVAV 1463

QY 326 --DET-----MFAIE----- 334

DB 1464 KANYKDRSSRVQALHHLKMIYESPSNITSINIYRRKQLLIIFYDIOAKYQDQDYGTLTIQ 1523

QY 335 -----DSIK--H-----NYDV 343

DB 1524 YASNAHNNTNAEVRLLKVERDYWINAKLSEQPKLLQLEITHMOKIRDVHIOVGLLVNDK 1583

QY 344 -----VSASSGFTG--TGLVGE 358

DB 1584 RKELESLBKMDANDPQSRGLLAEVNSPGTKHYDGNLMITTPERTIHFGFNSFTG--GP 1644

QY 359 KY-----W--QAIRALRRAGI-----P-----MVVATGNY 381

DB 1642 KYFGKVASHSINIEVEFEYBAGILPGLTHNWKAEILRTPFGDMRVNSLDAGIYSLKNL 1701

QY 382 ATSASSSSW-----DLVAN-----H--LKM 400

DB 1702 ILVNSTLEWADQKLQVGYKSDYDVNDQLSFDPVRFGINSTIRDIPTINVKVKMMDVKK 1761

QY 401 TDT-----G-NVT-----R-----TAH--- 412

DB 1762 VDTLEYLYGSGQNDTPNTYSMDSSWEIEKNORYNNYSGLVLVSPFKGYEKGGLVAHSL 1821

QY 413 EDALAVASAKQVTEPD-----KVN-----G--- 434

DB 1822 SDQRVSGA--ASINFLDREFTLTNGNVKKFTDNMLTVNITPLEKGTINARGLNEK 1879

QY 435 -----G-----SEFKRNIGAFPDKSI----- 452

DB 1880 KRHAVAEVRAPTAALGVEVLADIKNLLNFDVKSVAATPIESFQQAALFALNPERVDMRG 1939

QY 453 ----- 452

Db 1940 LWNVTGLGTGVHMQNITDPEYSYHVTPLAGPENGFIQVLLKKEFVFQFHGKMSNY 1999
Qy 453 -----TNNEDGT-----KAPSKUK-----FVY----- 469
Db 2000 KLGKINGEPKSDLVNLQSGNWELEMLYDADFKNLNAETDYKPDADBEYFSYFTNFQVD 2059
Qy 470 -----TCKGQODLIGLDL-----RGKIAMVDRIYTKDL----- 498
Db 2060 TLWPTFVGNVDIIEIDPVLVGVHVELPQGVKFEKRLUHPDYINVHNLTLVTPPFAVA 2119
Qy 499 -----KNAF-----KKAMD----- 507
Db 2120 KNIKSIHVHDLNFAFVERVKFIVNDKDKTQELGFVFNFTALQDNVKKPKAHVDQVTL 2179
Qy 508 -----KG----- 509
Db 2180 LTPYEMLEHYVGHIELDNAYKGNISAVTAHTLSMAASIENEDNFLTSGVQLETD 2239
Qy 510 -----ARAI-----MVVN----- 517
Db 2240 ALPHYGCQVYFKDPSAVDKAIDIRFEVTDNGLNQLHISTDWTDPSPYVNAVGRITKT 2299
Qy 518 -----TVNYNRDN-----WTEL----- 530
Db 2300 MLPLQWASTVLVIOQNPHLNFDLNLLSQGSTAYGARANKKXDVFNIEWTPMKNFR 2359
Qy 531 ----- 530
Db 2360 NISMHGTAIRSPDGRVDSVGLFYRNMATYEVTVGAVRMTNSLPIDVVLVQPKAGGRDG 2419
Qy 531 -----PAM-----CY-----EAD-----EGTKSOV-----F 546
Db 2420 VIELNIHEAGPKIRFSAIEDGKMKQMSGYSVSKTNGAMDVSVLVESTEPEIARINF 2479
Qy 547 -----SISGD-----DGVL----- 556
Db 2480 YGNLSPNSEGLVGLDLSLETPWKALGIDTVLHLSVDGLFKGHHVGEYKIGQYIGRSC 2539
Qy 557 ----- 559
Db 2540 LWSWILAEDMQLVLENLERNPAKPRIVHASAKYQNPQTFTLQAGGRLSVDSKWNLDV 2599
Qy 560 ----- 567
Db 2600 NGSAYKSVDDPKFRVITAPLPVGRHOLSAYQGNVISQGFNPDPFVLEASYESFEAQ 2659
Qy 568 -----VKR-----NN-----KEDFKDLKQ----- 582
Db 2660 NKLLSRISYKNATNNLKGHVEWGGKIQNLVVEGDFELLHKQAGREFSAKIITPKPKN 2719
Qy 583 -----Y-----YP-----IDME-----SFNSNKP-----NVG 599
Db 2720 EHTFALTGSYDLEKSHHNVGSLDPASRRITDLDSVSSLSNMHGFNSTLPTFLNV- 2778
Qy 600 DEKEDKFA-----P-----DTP----- 613
Db 2779 SWLKTDFTNNGKSYRYCRFCFWPDQDTAYFKLNSNYDSSNFHNLNGVEIEVPLAT 2838
Qy 614 ----- 623
Db 2839 RHRADIVYQKERNQDAGNVKVVYNEKQVLDGKYKRLQAKAPIYKETTDSLSENVKP 2898
Qy 624 -----PAGS----- 627
Db 2899 LGIHVSTRDASDPAGSQDKHIEIYELNQTQFNLTGELHSRATLKAQDFKVAIHPNR 2958
Qy 628 -----TSW----- 630
Db 2959 AVLSTKYEDVSEFVVRHHSKLELSETAWIGYNLELGNFSKVGNSQSFPALBIFPKNKL 3018
Qy 631 ----- 633
Db 3019 SSSGQYYMTDTNFNSDLSFQWLGNGYDQPKIHSNLQWKAEPHLRGDRHRTIALTVAH 3078

Qy 634 -----IDLLKPP-----VSAP 645
Db 3079 PLLSEKDNCKATYYRGLRDLRLTHLTIDYSEYDQLIELGQLTDRYSELGHTNTVTFHY 3138
Qy 646 GKNIKSTLNV-ING-----KST-----YG 663
Db 3139 GKHTASELDVOLNGTLAAMNSYYKTESTAHYKRDIFPARYG 3179
RESULT 15
AAU05396
ID AAU05396 standard; Protein; 26926 AA.
XX AC AAU05396;
XX DT 24-OCT-2001 (first entry)
XX DE Human titin (connectin) protein sequence.
XX KW Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
KW titin-related disease; zebrafish; heart failure; heart disease.
XX OS Homo sapiens.
XX PN WO200151666-A1.
XX PD 19-JUL-2001.
XX PF 12-JAN-2001; 2001WO-US01212.
XX PR 12-JAN-2000; 2000US-0175787.
XX PA (GEHO) GEN HOSPITAL CORP.
XX PI Fishman MC;
XX DR WPI; 2001-451869/48.
XX DR N-PSDB; AAS05390.
PT Determining if a subject has or is at risk of developing a
PT titin-related disease or condition, particularly heart failures,
PT comprises detecting the presence of a mutation in the titin gene -
XX Disclosure; Page 57-111; 114pp; English.
XX The present sequence representing human titin (also known as connectin)
is described in an invention relating to a novel method for determining
CC whether a subject has or is at risk of developing a titin-related
CC disease or condition. The method comprises analysing a nucleic acid of
CC a sample from the subject and detecting the presence of a mutation
CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
CC titin gene, which indicates that the subject has or is at risk of
CC developing a titin-related disease. The zebrafish which has a phenotype
CC similar to mammalian heart failure is used as a model. The method is
CC useful for detecting an increased likelihood of heart disease, such as
CC heart failure, in a patient, so that appropriate intervention can be
CC instituted before any symptoms occur. The method may also be used to
CC facilitate determination of etiology of an existing heart condition,
CC such as heart failure, to identify compounds that can be used to treat
CC or prevent heart conditions, in prenatal genetic screening, e.g. to
CC identify parents who may be carriers of a recessive titin mutation.
CC Compounds identified using the methods may be used to treat patients
CC that have or are at risk of developing heart disease, e.g. heart
CC failure.
XX Sequence 26926 AA;
SQ

Query Match 26.4%; Score 1161.5; DB 22; Length 26926;
Best Local Similarity 8.0%; Pred. No. 0.00016;
Matches 323; Conservative 169; Mismatches 164; Indels 3365; Gaps 140;
Qy 2 P---VVLAD-----TSSSE 12

Dd	3809	PETQAVLSDTEKIFPSAMSIEDINLTVLPKTLIAEPBGNYPOSSIEPPMHSHYLTLSVAE	3868
Oy	13	DAL-----NISD-----KEKVAENK-----EKH-ENI-----	33
Dd	3869	EVLSLIKETVSDTNEQGVTLQKQAGASALLISQSLAGHESLOSPOVMISQVNVPELV	3928
Oy	34	--HSAME-----TSOD-----FK--EKKTAVIEK-----	54
Dd	3929	PSEHCTEGKILIESANPLENAGODSAVRIEKGSLFRPLALEKQVLTKEEHSDNYVM	3988
Oy	55	-----EVSKN	60
Dd	3989	PPDOIIESKREPVAIKKVOEVRDLKSLSLSGIPBEORLNLKIQICRALQAANVASEQ	4048
Oy	61	P-----VIDNNTSN-----	69
Dd	4049	PGLPSEMLRNIEKVEAENVITQEPRIHMCWLYTSAKSYTEEVTIIEEDVDPOANLKM	4108
Oy	70	-----EAKI-----	74
Dd	4109	ELRDLALCAIYEEIDILTAEGPRIQGAKTSLOEEMDSFGSQKVEPIITEPESEKYLIS	4168
Oy	75	-----KEENSNS-----	82
Dd	4169	TEEVSYFNVQSRVKYLDATPVTKGVAASVSDKODESLKSEEEKESBSSSESGTEEVAIV	4228
Oy	83	-----OD-----YTDSFVN-----KNTENPKKE-----	101
Dd	4229	KIQEAGGLIKEDGSMITHPLVDVITSEBGDIYHLTTSITNAKEVWYFENKLVPSDEKFK	4288
Oy	102	-----DKVV-----YIAE-----	109
Dd	4289	CLQDQNTYTLVIDKVNTEDHQGEYVCEALNDGSKTATSAKTLTVKRAAPVIRKRIEPLV	4348
Oy	110	-----FK-----DKES-----GE-----	117
Dd	4349	ALGHLAKETCEIQASPNVRPWFQKAGREIYESDKSIRSKYISSLEILRTQVVDGGEYT	4408
Oy	118	--KAKELSLKNKTYL-----	132
Dd	4409	CKASNEYSVSGCTATLTVTPGEGEKVKRLPERKPEPKEEVLKSVLRKPEEPEPVE	4468
Oy	133	-----	132
Dd	4469	PKLEKVKKPAVPEPPPKPVEVEVPVTKREKRIPEPTKVPEIKPALPLPAPERKRP	4528
Oy	133	-----	132
Dd	4529	EAENVKTIKPPVEPEPTPIAAPVTVVVKKAELAKAPEAAKPKGPIKGVPKTTPSIE	4588
Oy	133	-----YTY-----D-----RF-----	139
Dd	4589	AERRKLKRGSGGKEPPDEAPFTYQKAVPLKFEVKEIKOIIITSESEFVSALFECVLSPS	4648
Oy	140	-----NGSAIEFTTP-----DNLDK-----	153
Dd	4649	TAITTMWDGSIIRSPKRFIADGDKRKLHIIDVQSDAGEYTCVLGNKEKSTAKL	4708
Oy	154	-----IKOIE-----	160
Dd	4709	VVEELPVRFVKTLBEEVTVKQPLYLSCELNKERDVVVRKDGKIVKEKGRIVGVIGL	4768
Oy	161	-----SSVP-----	172
Dd	4769	MKALTINDADTDCTVTVTVENANNLESCSCVKVVEVIRDWLVKPIRDQHVKPGTAIF	4828
Oy	173	-----	175
Dd	4829	ACDIAKDPNIIKMFQGYDEIAPRPDKTEILRDGSHLYLKIGNAMPEDIAEVAVEIEGR	4888
Oy	176	-----ARK-----	178

Dd	4889	YPAKLTIGERVELLKPLEDVITYEKESASFDAEISENDIDGOMKLGCELLRPSPTCEIK	4948
Qy	179	-EIG-----VEE-AIDY--LKSINAP-----F	196
Dd	4949	AEGGRFLTLHKVLDQAGEVLVYQALNATITALLTVEIKELDFAPVLKDVTPBRQRAF	5008
Qy	197	-----GKN-----F-----DGRGVNISNIDTGD-----	215
Dd	5009	ECVLTRBANVIMSGPDIIKSSDPFDIIADGKHILVJNDSQFDDGCVTAVEGKKTSA	5068
Qy	216	-----YRKAMRI-----	223
Dd	5069	RLFVGTGIRLKMSPLEDOTVEGETATFVCELSHEKQHVWFKQDAKLHSTRVLLISSEG	5128
Qy	224	-----DD-----	225
Dd	5129	KTHKLEMEKVTLDDISOIKAQVKELSSTAQVLAEADPYFTVKLHDKTAVBKEITLKCE	5188
Qy	226	---DAKASMF-----KKEDLK-----GTDKN-----	244
Dd	5189	VSKDVPV-KWFKDGEELVPSPKYSIKADGLRILIKIKAKDLKDGVEVCDGCTKTYANV	5247
Qy	245	-----Y-----W-----LSDKIP	252
Dd	5248	TVEARLIEVEKPLVGEVEFVGETAHFELIESEPVGHWKUKGQPLTASPDCIIEBGKK	5307
Qy	253	HAFNYN--G--GKITEV-----KYDD-----	270
Dd	5308	HILIHNCQGLMGTEBVSFOAANAKSAAHLKYLPLFITPLSDVKVPEKBEAKPECEVS	5366
Qy	271	-----GRDYF-----D--PHGM-----HIAG--IL	286
Dd	5368	REPKTFWLKGTQETGDDRFELIKDGTKSHMVIKSAAFEDEAKYMFADBDKHTSGKLI	5427
Qy	287	AG-----ND-----TEOD-----	294
Dd	5428	EGIRLKLFTPLKDVTAKEKESAVFTVELSHDNIRVKWFKNDQRLHTRTSVMODEGKTHS	5487
Qy	295	-----IKNFNGI-----	301
Dd	5488	ITPFDLSIDTTSQIRVEAMGMSSEAKLTVLEGDYPFGKLDQDYTVEGDEVILOCELSKA	5547
Qy	302	---DG--IAP--NA-----QIFSYKM-----YS-DAGSGFAGDETM--	329
Dd	5548	DAPVKWFDGKEIKPSKNAVIKTDGKKRMLLKALKALSIDIGQYCTDGT---DXTSKL	5603
Qy	330	-----FNAI-----EDSI-----	337
Dd	5604	DIEDREIKVLPRLSHVEVMEETETARFETEISEDDIHANWKLKGALLQTPDCEIKEEGKI	5663
Qy	338	---KHN-----VDV-----VSVSSG-----	349
Dd	5664	HSVLVHNCRLDQGTGVDFOAANVKSSAHLRVKPRVIGLRLPKDVTYAGETAIPDCELS	5723
Qy	350	-----	349
Dd	5724	YEDIPVEWYVLKGGKLEBSPDKVVPVRSSEKQVHTLTLRDVCLBDAGEVQULAKDFKTHANLV	5783
Qy	350	---FT-----GT--GLVGEKY--	360
Dd	5784	KEPVEFETKPLEDOTVEGATAVLECEVSRENAKYKWFKNQTEILKSKKYEIVADGRVVK	5843
Qy	361	-----	360
Dd	5844	LVIHDCIPEDIKYTCAKDFKTCSCNLNVPPHVEFLRPLTDLQVREKEMARFECESLRE	5903
Qy	361	---W-----QAIRLKRAGIPW--	375
Dd	5904	NAKVKWFDGAEIKKKKKYDIISKGAVRILVINKCLDDDEABVSCVARTATSGMLTVLE	5963
Qy	376	---VATGVNATSASSS-----W--D-----L	392
Dd	5964	BEAFTFKNLNIEVSEDTGIKLVCEVSPGAEVIMYKQDEBIEITGRYEILTEGRKRILV	6023

QY 393 VANNHLMKTDGTVN-----TRT-----AAHE----- 413
Db 6024 IONAHLE--DAGNYNCRPLSSRTDGVKVKVHELAAEFISKPONLEILEGEKAEFVCSISKE 6081
QY 414 -----DATA-----VASA----- 421
Db 6082 SFPVQWKRDDKLTESGDKYDVITADGKRVLVVKDATLQDMGTYYVMVGAARAAHLTVIE 6141
QY 422 -----KNQTVFED-KYNIGGESPKY-RNIGAPFDKSK----- 451
Db 6142 KLRIVVPLKDRVKEQEVVFNCEVNTGAKAKWFERNEEAIIPDSSKIILQKDLVYTLRI 6201
QY 452 -----ITNED----- 457
Db 6202 RDAHLDQANVNSLTNRHGENVKSAANLIVEEDLRIVEPLKDIETWEKKSVTFWCKVN 6261
QY 458 -----GKAPSKLKFFVY-----IGKQD----- 475
Db 6262 RLNVTLKWTNGEEVFPFDRVSRYVDKYKMLTIKDCGPPDEGEIYTAGQDKSVAELLI 6321
QY 476 ----- 475
Db 6322 IEAPTEFVEHLEDQTVTEFDDAVFSCQLSREKANVKWYRNGREIKEGKKYKPEKDGSIHR 6381
QY 476 -----QDLI-----GLD----- 482
Db 6382 LIIKCDRLDDECEYACGVEDRSRARLFVEEIPVEIIRPPQDILEAPGADVVFVLAELNKO 6441
QY 483 -----LR-----GKI-----AVMD-----RYTKD----- 497
Db 6442 KVEVOWLRNNMVVQGDKHOMSEKGIHLRLOICDIKPRDQGEYRFIAKDKKEARAKLELAA 6501
QY 498 -----LKNA-----FK----- 503
Db 6502 APKIKTADQDLVDVGKPLTMVVPYDAYPKAEAEWFKENEPLSTKTIDTTABQTSFRILE 6561
QY 504 -KAMDKG----- 509
Db 6562 ARKGDGKRYKIVLONKHGAEGFINLKVIDVPGPVNLEVTETFDGEVSLAWEBPLTDGG 6621
QY 510 ----- 509
Db 6622 SKIIGVYVVERDIKRTWTVLATDRAESCEFTVTGLQKGVEYLFVRSARNRVGTGEVET 6681
QY 510 -----AR-----AIMVNV-----TVNY----- 521
Db 6682 DNPVEARSKYDVGPPPLNVITIDVNRFGVSLTWEPPEYDGGAEITNYVIELRDKTSIRWD 6741
QY 522 -----Y-----NR----- 524
Db 6742 TMTVRAEDLSATVTDVVEGQESYFRVRAQNRIGVGKPSAATPFVKVADPIERPSPVNL 6801
QY 525 -----DNW-----TEL----- 530
Db 6802 TSSDOTQSSVQLKWEPPPLKDGSPILGYIIERCEBCKDNWIRCNWKLVPETYKYVTGLEK 6861
QY 531 ----- 532
Db 6862 GNKYLVRSAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEIVVNPNTILVPS 6921
QY 533 MGYE-----ADEG-----TKSQVPSISG 550
Db 6922 TGYPRPTATWCFQDKVLETGDRVMMKTL SAYAELVISPSRSDKGIYTLKLENRVKTIISG 6981
QY 551 -----DDG----- 559
Db 6982 EIDVNVIAEPSAPKELKFGDITKDSVHLTWEPDDDDGSGPLTGYVVVEKREVSARKTWM 7041
QY 560 -----IN-----PDKKTEYK----- 569
Db 7042 DFVTDLEFTVPLVQGEYLFKVCARNKCGPGEPAVVDSPVNMSTPATVDPDENVKWRD 7101

QY 570 RNNKEDF-----KDK-----LEQ--YY- 584
Db 7102 RTANSIFLTWDPFKNDGSGSRIGKIVERCPRGSDKWACGEPVAETKMEVTLGEBGKMYA 7161
QY 585 -----P-----IDM-----ESF----- 591
Db 7162 YRVKTLNRQAGSKPSRPTETIQAVDTOBAPBIFLDVKLLAGLTVKAGTKIELPATVTKP 7221
QY 592 -----NSNK----- 596
Db 7222 EPKITWTKADMILKODKRITITENVPKSTVTIVDSKRSDDTGYIIEAVNVCGRATAVVEV 7281
QY 597 NVGDB-----K 602
Db 7282 NVLDKPGPPAAPDITDVTNESCLLTWNPPRDDGSGKITNYVVERRATDSEVVMHKLSSIVK 7341
QY 603 BIDFK----- 607
Db 7342 DTNFKATKLI PNKEYIFRVAENMYGAGEPVQASPIAKYQFDPGPPPTLRLEPSDITKDA 7401
QY 608 -----FAPDPTDK----- 614
Db 7402 VLTWCEPDDDDGSGSPITGYWVERLDPDPTDKWVR CNKMPVKDITTVRVKGLTNKKKYRFRVL 7461
QY 615 -----ELYKEDIIVP----- 624
Db 7462 AENLAGCPKPSKSTEPILIKDIPDPWPFGPKTVKDVGKTSVRLNWTYPBHDGGAKIESY 7521
QY 625 -----AGTSW-----G----- 631
Db 7522 VIEMKTGTDSEWVRVAEGVPTTQHLLPGLMEGQESYFRVRAVKNKAGESESPSPDVLCR 7581
QY 632 -----PR-----IDLLK 639
Db 7582 EKLYPSPPRWLEVINITKNTADLKWTVPEKDGSGSPITNYIVERKDRVRKGMQTVDTTVK 7641
QY 640 -----P-----DVS-- 643
Db 7642 DTKCTVTPLTBGSLYVFRVAENAIGQSDYTEIEDSVLAKDTFTTPGPPYALAVVDVTKR 7701
QY 644 -----APGKN-----IKS-----TLNVI----- 656
Db 7702 HVDLKWEPPKNDGGRPIQRYVIEKRLGTRWVKAGTAGPCNFRVTDVIEGTEVQFOV 7761
QY 657 -----NGKS--TVG 663
Db 7762 RAENAGVGHPSPEPTTEILSIEDPTSPSPPLDLHLVTDAGRKHIAIAWKPPKNGGSPIG 7821
QY 664 Y 664
Db 7822 Y 7822

Search completed: October 9, 2003, 15:50:54
Job time : 127 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2003, 15:55:23 ; Search time 47 Seconds

(without alignments)
2242.435 Million cell updates/sec

Title: US-09-590-991-6

Perfect score: 4398
Sequence: 1 YPVVLADTSSSEDAALNISDK.....PGKNIKSTLVINCKSTRGY 664Scoring table: BLOSUM30
Gapop 1.0 , Gapext 0.1

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4398	100.0	664	22 AAB48342	S. pneumoniae Sp12
2	4398	100.0	2120	21 AAY81710	Streptococcus pne
3	4395.7	99.9	2140	24 ABU01020	S. pneumoniae type
4	4171	94.8	1007	19 AAW61246	Streptococcus pne
5	4171	94.8	1007	23 ABP54664	S. pneumoniae Sp12
6	3114.7	70.6	26926	22 AAU05396	Human ctitin (conn
7	3104.2	70.6	31267	24 ABG74786	Human RGS11 protei
8	3098.2	70.4	7107	22 ABB58144	Drosophila melanog
9	3070.2	69.8	5701	22 ABB51536	Human liver peptid

10	3070.2	69.8	5701	22 ABB36684	Peptide #4190 enco
11	3070.2	69.8	5701	22 ABB22021	Protein #4020 enco
12	3070.2	69.8	5701	22 AAM57448	Human brain expres
13	3070.2	69.8	5701	22 AAM69843	Human bone marrow
14	3070.2	69.8	5701	22 AAM05328	Peptide #4010 enco
15	3058.8	69.5	6619	22 ABBG33329	Novel human diagno
16	3056.9	69.5	15281	15 AAR44929	T. niueum Cyclospo
17	3032.8	69.0	4643	22 ABB71609	Drosophila melanog
18	3030.6	68.9	7718	22 ABBG1181	Novel human diagno
19	3029.6	68.9	4345	22 ABBB6417	Drosophila melanog
20	3028.2	68.9	7201	22 ABB71136	Drosophila melanog
21	3026	68.8	4796	22 ABBB5865	Drosophila melanog
22	3020.2	68.7	5909	22 ABBG3295	Novel human diagno
23	3019.1	68.6	6077	24 ABBP5996	White spot syndrom
24	3017.5	68.6	5635	23 ABBP6091	Novel human protei
25	3016.9	68.6	4689	22 ABBG4550	Novel human diagno
26	3016.6	68.6	4636	22 AABE1197	Human COL16 (or C5
27	3015.1	68.6	6025	22 AAG84939	Shrimp white spot
28	3015	68.6	6025	22 AAG84939	Human kidney calci
29	3015	68.6	4655	17 AAR97210	Human parathyroid
30	3015	68.6	4655	19 AAW43313	Human parathyroid
31	3015	68.6	4655	19 AAW43314	Human LRP1B protei
32	3013.5	68.5	4599	24 ABB56837	Human calcitum sens
33	3013	68.5	4655	17 AAR97208	Human calcitum sens
34	3011	68.5	4655	19 AAW43311	Human calcitum sens
35	3011	68.5	4655	24 ABB56836	Human megalin prot
36	3010.6	68.5	5464	22 ABBG11810	Novel human diagno
37	3010	68.4	4655	17 AAR97209	Human placental ca
38	3010	68.4	4655	19 AAW43312	Human placental ca
39	3000.5	68.2	5107	22 ABBB5257	Drosophila melanog
40	2999.7	68.2	4472	22 ABBB60101	Drosophila melanog
41	2995.6	68.1	4555	23 AAM52106	Rat fat 3 protein
42	2987	67.9	5147	22 ABB59831	Drosophila melanog
43	2979.4	67.7	4741	22 AAUJ1612	Novel human secret
44	2972.7	67.6	4349	23 AAUJ9940	Human protocadheri
45	2972.7	67.6	4349	23 ABB97540	Novel human protei

ALIGNMENTS

RESULT 1
ID AAB48342 standard; Protein: 664 AA.
XX AAB48342;
AC
XX
XX
20-APR-2001 (first entry)
DT
XX
XX
S. pneumoniae Sp128 polypeptide.
XX
XX
Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW bronchial; lung; blood; infection; immune response; immunotherapy;
KV antibacterial; auditory; vaccine.
XX
XX
OS Streptococcus pneumoniae.
XX
XX
WO200076540-A2.
XX
XX
21-DEC-2000.
XX
XX
PD
XX
XX
09-JUN-2000; 2000WO-US15925.
XX
XX
PF
XX
XX
10-JUN-1999; 99US-0138453.
XX
XX
PA (MEDI-) MED IMMUNE INC.
XX
XX
PI Adamou JE, Choi GH;
XX
XX
WPI; 2001-112197/12.
DR
XX
XX
N-PSDB; AAC84741.
XX
XX
PT New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
PT blood infections -
XX
XX
XX Claim 4; Page 47-50; 54pp; English.
XX
XX The invention relates to novel immunogenic polypeptides, Spi28 and Spi30
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC the treatment and prevention of pneumococcal infections, particularly
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC bronchial, lung or blood infections. The antigens are used as immunogenic
CC agents to stimulate an immune response. The antisera and antibodies may
CC also be used in diagnosing and treating pneumococcal infections.
CC Recombinant polypeptides serve as a mechanism for stimulating production
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC as reagents in other processes such as affinity chromatography. The
CC present sequence represents the S. pneumoniae Spi28 polypeptide.
XX
XX Sequence 664 AA;
SQ

Query Match 100.0%; Score 4398; DB 22; Length 664;
Best Local Similarity 100.0%; Pred. No. 7.9e-77;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YPVVLADTSSSEDALNISDKKEVAENKEKHENIHSAMETSQDFKEKKTAVIKEVVSXN 60
DB 1 YPVVLADTSSSEDALNISDKKEVAENKEKHENIHSAMETSQDFKEKKTAVIKEVVSXN 60
QY 61 PVIDNNTSNEAKIKEENSNSQGDYTDTSFVNKNTENPKEDKVYIAEFKDKESGEKAI 120
DB 61 PVIDNNTSNEAKIKEENSNSQGDYTDTSFVNKNTENPKEDKVYIAEFKDKESGEKAI 120
QY 121 KELSSLKNTKVLTYDYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKQVPMNHARKEI 180
DB 121 KELSSLKNTKVLTYDYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKQVPMNHARKEI 180
QY 181 GVEEADYLSINAPFGKNGFGRGMVINSIDTGDYRHKAMRIIDDKASMRFKEDLKG 240
DB 181 GVEEADYLSINAPFGKNGFGRGMVINSIDTGDYRHKAMRIIDDKASMRFKEDLKG 240
QY 241 TDKNYWLSDKIPAFNYNGGKIIVKDYDDGRDYFDPHGMIAGILAGNDTEQDIKNFNG 300
DB 241 TDKNYWLSDKIPAFNYNGGKIIVKDYDDGRDYFDPHGMIAGILAGNDTEQDIKNFNG 300
QY 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVVDVSVSGFTGTGLVGEKY 360
DB 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVVDVSVSGFTGTGLVGEKY 360
QY 361 WQAIRALKAGIPMVATGNVYATSSSSWDLVANNHLKMTDTGNVTRTAHEDAIAS 420
DB 361 WQAIRALKAGIPMVATGNVYATSSSSWDLVANNHLKMTDTGNVTRTAHEDAIAS 420
QY 421 AKQTVDFKVNIGGESFKYRNIGAFDPKSKITTNEDGTAKPSLKFVYIGKQDQDLIG 480
DB 421 AKQTVDFKVNIGGESFKYRNIGAFDPKSKITTNEDGTAKPSLKFVYIGKQDQDLIG 480
QY 481 LDLRGKIAVMDRIYTKOLKNAFKKAMDKGARAIMVNTVNYNRDNMTLPAWGVEADEG 540
DB 481 LDLRGKIAVMDRIYTKOLKNAFKKAMDKGARAIMVNTVNYNRDNMTLPAWGVEADEG 540
QY 541 TKSOVFSISGDDGVKLNMINPDKTEVKRNKNEKDFDKLQYYPIDMESFNSKNPNVGD 600
DB 541 TKSOVFSISGDDGVKLNMINPDKTEVKRNKNEKDFDKLQYYPIDMESFNSKNPNVGD 600
QY 601 EKEIDFKFAPDTEKELYKEDIIVPAGSTSWGPRIDLLKPDVSPAGPKNIKSTLVNINGS 660
DB 601 EKEIDFKFAPDTEKELYKEDIIVPAGSTSWGPRIDLLKPDVSPAGPKNIKSTLVNINGS 660
QY 661 TYGY 664
DB 661 TYGY 664

RESULT 2
RAY81710
ID AAY81710 standard; Protein; 2120 AA.
XX
XX AAY81710;
XX
XX 02-JUN-2000 (first entry)
XX Streptococcus pneumoniae protein sequence ID3.
XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
XX Streptococcus pneumoniae.
XX
XX WO200006738-A2.
XX
XX 10-FEB-2000.
XX
XX 27-JUL-1999; 99WO-GB02452.
XX
XX 27-JUL-1998; 98GB-0016336.
XX
XX 19-MAR-1999; 99US-0125329.
XX
XX (MTCR-) MICROBIAL TECHNIQS LTD.
XX
XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX
XX WPI: 2000-195301/17.
XX
XX N-FSDB; AAZ91806.
XX
XX Streptococcal proteins and polynucleotides useful for diagnosis,
XX treatment and prophylaxis of bacterial infections -
XX
XX Claim 2; Page 41-42; 76pp; English.
XX
XX This sequence represents a Streptococcus pneumoniae protein of the
XX invention. The proteins (or their homologues, derivatives and/or
XX fragments) are useful as immunogens or antigens. Immunogenic or antigenic
XX compositions comprising the proteins are useful as vaccines and also in
XX diagnostic assays. The sequences are useful for the detection or
XX diagnosis of S. pneumoniae infection, by contacting a sample to be tested
XX with them. Agents capable of antagonising, inhibiting or interfering with
XX the function or expression of the protein or polypeptide are useful in
XX medical compositions in the treatment or prophylaxis of S. pneumoniae
XX infection. As the sequences can be used to treat S. pneumoniae infection,
XX they can be used to treat bacterial pneumonia, which has high rates in
XX young children, the elderly, and in patients with predisposing conditions
XX such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
XX or with immunosuppressive disorders, especially AIDS. They can also be
XX used to treat pneumococcal septicaemia, otitis media, sinusitis, and
XX meningitis.
XX
XX Sequence 2120 AA;

Query Match 100.0%; Score 4398; DB 21; Length 2120;
Best Local Similarity 100.0%; Pred. No. 4.8e-74;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YPVVLADTSSSEDALNISDKKEVAENKEKHENIHSAMETSQDFKEKKTAVIKEVVSXN 60
DB 1 YPVVLADTSSSEDALNISDKKEVAENKEKHENIHSAMETSQDFKEKKTAVIKEVVSXN 60
QY 61 PVIDNNTSNEAKIKEENSNSQGDYTDTSFVNKNTENPKEDKVYIAEFKDKESGEKAI 120
DB 61 PVIDNNTSNEAKIKEENSNSQGDYTDTSFVNKNTENPKEDKVYIAEFKDKESGEKAI 120
QY 121 KELSSLKNTKVLTYDYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKQVPMNHARKEI 180
DB 121 KELSSLKNTKVLTYDYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKQVPMNHARKEI 180

QY 181 GVEBAIDYLSKINAPFGKIPDGRGVISNIDTGTDRHKAMRIDDAKASNRFKKEDLKG 240
DB 181 GVEBAIDYLSKINAPFGKIPDGRGVISNIDTGTDRHKAMRIDDAKASNRFKKEDLKG 240
QY 241 TDKNYWLSDKIPHAFNYNGGKITVEKIDGRDYPDPGHMIIAGIAGNDTEODIKNFNG 300
DB 241 TDKNYWLSDKIPHAFNYNGGKITVEKIDGRDYPDPGHMIIAGIAGNDTEODIKNFNG 300
QY 301 IDGIAIPNAQIFSYKMYSDAGSGFAGDETFPHAIEDSIKHNDVVSVSSTGFTGLVGEKY 360
DB 301 IDGIAIPNAQIFSYKMYSDAGSGFAGDETFPHAIEDSIKHNDVVSVSSTGFTGLVGEKY 360
QY 361 WQAIRALRKAGIPMVVATGNVATSSSSSWDLVANNHAKMTDGTGVNRTTAHEDAIAS 420
DB 361 WQAIRALRKAGIPMVVATGNVATSSSSSWDLVANNHAKMTDGTGVNRTTAHEDAIAS 420
QY 421 AKNOTVEFDKVNIGESFKYRNIGAFDPKSKITTNEDGTAKPSKLKPYIIGKODQDILIG 480
DB 421 AKNOTVEFDKVNIGESFKYRNIGAFDPKSKITTNEDGTAKPSKLKPYIIGKODQDILIG 480
QY 481 LDLRGKIAVMDBRITTKDKNAFKKAMDGARAIVNVVNVYNDNMTELPMAGYEADG 540
DB 481 LDLRGKIAVMDBRITTKDKNAFKKAMDGARAIVNVVNVYNDNMTELPMAGYEADG 540
QY 541 TKSQVFSISGDDGVKLMNINPDKTEVKRNKKEDFKDLKQYYPIDMESFNSKNPNVGD 600
DB 541 TKSQVFSISGDDGVKLMNINPDKTEVKRNKKEDFKDLKQYYPIDMESFNSKNPNVGD 600
QY 601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIIDLLKPDVSAFGKNIKSTLVNINGS 660
DB 601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIIDLLKPDVSAFGKNIKSTLVNINGS 660
QY 661 TYGY 664
DB 661 TYGY 664

RESULT 3
ABU01020
ID ABU01020 standard; Proteoin; 2140 AA.

XX AC ABU01020;
XX DT 11-FEB-2003 (first entry)
XX DE S. pneumoniae type 4 strain protein from coding region #590.
XX DB Bacterial meningitis; pneumonia; sepsis; otitis media;
XX KM ear infection; antiinflammatory; antibacterial; immunostimulant;
XX KM auditory; respiratory; gene therapy; vaccine.
XX OS Streptococcus pneumoniae type 4 strain.
XX PN WQ200277021-A2.
XX PD 03-OCT-2002.
XX PF 27-MAR-2002; 2002WO-IB02163.
XX PR 27-MAR-2001; 2001GB-0007658.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Maelignant V, Tecteljin H, Fraser C;
XX DR WPI; 2003-040579/03.
XX DR N-PSDB; ABX06302.
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection

XX XX
XX Claim 1; SEQ ID No 1180; 56pp; English.
PS The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB856654. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2469 proteins expressed by the identified coding regions from the
CC genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 2140 AA;

Query Match 99.9%; Score 4395.7; DB 24; Length 2140;
Best Local Similarity 99.6%; Pred. No. 5.6e-74;
Matches 664; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

QY 1 Y-P--VVLADTSSSEDLNLSDEKEVAENKEKHENHSAMETSODPFEKKTAVIKEKEV 57
DB 19 YAPNEVVLADTSSSEDLNLSDEKEVAENKEKHENHSAMETSODPFEKKTAVIKEKEV 78
QY 58 SKNPVINDNNTSNEBAKIKENSNSKSGDYTDSFVNKKTENPKKEDKVVYIAEFKDSGE 117
DB 79 SKNPVINDNNTSNEBAKIKENSNSKSGDYTDSFVNKKTENPKKEDKVVYIAEFKDSGE 138
QY 118 KAIFELSLKNTKVLVYDRFNGSALETTPDNLDKIKOIEGSSVERAQVQPMNMHAR 177
DB 139 KAIFELSLKNTKVLVYDRFNGSALETTPDNLDKIKOIEGSSVERAQVQPMNMHAR 198
QY 178 KEIGVEBAIDYLSKINAPFGKIPDGRGVISNIDTGTDRHKAMRIDDAKASNRFKKED 237
DB 199 KEIGVEBAIDYLSKINAPFGKIPDGRGVISNIDTGTDRHKAMRIDDAKASNRFKKED 258
QY 238 LKGTDKNYWLSDKIPHAFNYNGGKITVEKIDGRDYPDPGHMIIAGIAGNDTEODIKN 297
DB 259 LKGTDKNYWLSDKIPHAFNYNGGKITVEKIDGRDYPDPGHMIIAGIAGNDTEODIKN 318
QY 298 FNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFPHAIEDSIKHNDVVSVSSTGFTGLV 357
DB 319 FNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFPHAIEDSIKHNDVVSVSSTGFTGLV 378
QY 358 EKYWQAIRALRKAGIPMVVATGNVATSSSSSWDLVANNHAKMTDGTGVNRTTAHEDAI 417
DB 379 EKYWQAIRALRKAGIPMVVATGNVATSSSSSWDLVANNHAKMTDGTGVNRTTAHEDAI 438
QY 418 VASAKNOTVEFDKVNIGESFKYRNIGAFDPKSKITTNEDGTAKPSKLKPYIIGKODQD 477
DB 439 VASAKNOTVEFDKVNIGESFKYRNIGAFDPKSKITTNEDGTAKPSKLKPYIIGKODQD 498

```

QY 478 LIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNTVYNNRDNTWTELPAMGYEA 537
DB 499 LIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNTVYNNRDNTWTELPAMGYEA 558
QY 538 DEGTQSVFISGGDGVKLMNINPDKTEVKRNKEDFKDKLEQYYPIDMESFNSKNPN 597
DB 559 DEGTQSVFISGGDGVKLMNINPDKTEVKRNKEDFKDKLEQYYPIDMESFNSKNPN 618
QY 598 VGDEKEIDFKPAPDTEKELYKEDIIVPAGTSWGPRIIDLLKPDVSAFGKNIKSTLNVIN 657
DB 619 VGDEKEIDFKPAPDTEKELYKEDIIVPAGTSWGPRIIDLLKPDVSAFGKNIKSTLNVIN 678
QY 658 GKSTYGY 664
DB 679 GKSTYGY 685

RESULT 4
AAW61246
ID AAW61246 standard; Protein; 1007 AA.
XX
XX AAW61246;
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SPI22 protein.
XX
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
XX WO9818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US19422.
XX
XX 31-OCT-1996; 96US-0029960.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX WPI; 1998-272224/24.
XX N-PSDB; AAV27431.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX pneumoniae - or their epitope-containing fragments, useful in
XX protective or therapeutic vaccines, and for diagnosis
XX
XX Claim 11; Page 92-93; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX can be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly to detect specific
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein are used to
XX detect corresponding antigens, to purify the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose.
XX
XX Sequence 1007 AA;
XX
XX Query Match 94.8%; Score 4171; DB 19; Length 1007;
XX Best Local Similarity 100.0%; Pred. No. 4.4e-71;
XX Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 38 ETSQDFKEKKTAVIKEVSVSKNPVIDNNTSNEBAKIKEENSKSQGDYTDTSFVNKNTEN 97
DB 1 ETSQDFKEKKTAVIKEVSVSKNPVIDNNTSNEBAKIKEENSKSQGDYTDTSFVNKNTEN 60
QY 98 PKKEDKVYIAEAFKDKESGEKAIKELSSLKNTKVLYTYDRIENGSAIETTPNLDKIKOI 157
DB 61 PKKEDKVYIAEAFKDKESGEKAIKELSSLKNTKVLYTYDRIENGSAIETTPNLDKIKOI 120
QY 158 EGISVSERAQVQPMNMNHARKEIGVEEADYDKSINAPFGKNFDRGMVVISNIDTGTDIR 217
DB 121 EGISVSERAQVQPMNMNHARKEIGVEEADYDKSINAPFGKNFDRGMVVISNIDTGTDIR 180
QY 218 HKAMRIDDDAKASMRFKKEDLKGTDKNYLSDKI PHAFNYNGGKITVVEKYDDGRDYFDP 277
DB 181 HKAMRIDDDAKASMRFKKEDLKGTDKNYLSDKI PHAFNYNGGKITVVEKYDDGRDYFDP 240
QY 278 HGMHIAGILAGNDTEQDIKNFNIGIDGIAPNAQIFSYKMYSDAGSGFAGDETWFHAIEDSI 337
DB 241 HGMHIAGILAGNDTEQDIKNFNIGIDGIAPNAQIFSYKMYSDAGSGFAGDETWFHAIEDSI 300
QY 338 KHNVDVSVSSGFTGTGLVGEKYWQAIIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 397
DB 301 KHNVDVSVSSGFTGTGLVGEKYWQAIIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 360
QY 398 LKMTDTGNTVTRTAHEDAIASAKNQTVEFDKVNIGGESPKYRNIGAFDFKSKITTND 457
DB 361 LKMTDTGNTVTRTAHEDAIASAKNQTVEFDKVNIGGESPKYRNIGAFDFKSKITTND 420
QY 458 GTKAPSKLKFVYIGKQDQDLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVYN 517
DB 421 GTKAPSKLKFVYIGKQDQDLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVYN 480
QY 518 TVNYYNRDNTWTELPAMGYEADGTSQVFSISGDDGVKLMNINPDKTEVKRNKEDFK 577
DB 481 TVNYYNRDNTWTELPAMGYEADGTSQVFSISGDDGVKLMNINPDKTEVKRNKEDFK 540
QY 578 DKLEQYYPIDMESFNSKNPNVGDEKEIDFKPAPDTEKELYKEDIIVPAGTSWGPRIIDLL 637
DB 541 DKLEQYYPIDMESFNSKNPNVGDEKEIDFKPAPDTEKELYKEDIIVPAGTSWGPRIIDLL 600
QY 638 LKPDVSAFGKNIKSTLNVINGKSTYGY 664
DB 601 LKPDVSAFGKNIKSTLNVINGKSTYGY 627

RESULT 5
ABP54664
ID ABP54664 standard; Protein; 1007 AA.
XX
XX AC ABP54664;
XX
XX DT 04-SEP-2002 (first entry)
XX
XX DE S. pneumoniae SPI22 protein sequence SEQ ID NO:216.
XX
XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection.
XX
XX OS Streptococcus pneumoniae.
XX
XX PN US2002061545-A1.
XX
XX PD 23-MAY-2002.
XX
XX PF 22-JAN-2001; 2001US-0765272.
XX
XX PR 30-OCT-1997; 97US-0961083.
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.

```


PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX
 PI Choi GH, Kunsch CA, Baraeh SC, Dillon PJ, Dougherty B, Fannon MR,
 PI Rosen CA,
 XX
 DR MPI: 2002-479261/51.
 DR N-PSDB; ABQ84899.
 XX
 PT New Streptococcus pneumoniae antigens, useful for detecting
 PT Streptococcus and for preventing or attenuating disease caused by
 PT Streptococcus infection -
 XX
 PS Claim 11, Page 50; 70pp; English.
 XX
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP94557 to ABP94669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
 CC which are used in an example from the present invention.
 CC
 XX
 SQ Sequence 1007 AA;
 Query Match 94.8%; Score 4171; DB 23; Length 1007;
 Best Local Similarity 100.0%; Pred. No. 4,4e-71;
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 ETSODPEKKRAVKEKEVSKNPVIDNNNTNEEAKIENSNKSGDYTDSFVNKNTEN 97
 DB 1 ETSODPEKKRAVKEKEVSKNPVIDNNNTNEEAKIENSNKSGDYTDSFVNKNTEN 60
 QY 98 PKKEDKVVYIAEFKDKESGEKAIKELSLKNTKVLTYTDRIENGSAIETTPDNLKIKOI 157
 DB 61 PKKEDKVVYIAEFKDKESGEKAIKELSLKNTKVLTYTDRIENGSAIETTPDNLKIKOI 120
 QY 158 EGISSEBAQKVQPMNHARKEIGVEAIDYLSKINAPFGKDPGRGVNISNIDTGYDR 217
 DB 121 EGISSEBAQKVQPMNHARKEIGVEAIDYLSKINAPFGKDPGRGVNISNIDTGYDR 180
 QY 218 HKARIDDDAKASRFKEDLKGTDKNWLSDKIPHANVYNGSKITVEKDDGRDYDP 277
 DB 181 HKARIDDDAKASRFKEDLKGTDKNWLSDKIPHANVYNGSKITVEKDDGRDYDP 240
 QY 278 HGMHIAAGLADNTEODIKNFNGIDGAPNAQIFSYKMYSDAGSGFAGDETFPAIEDSI 337
 DB 241 HGMHIAAGLADNTEODIKNFNGIDGAPNAQIFSYKMYSDAGSGFAGDETFPAIEDSI 300
 QY 338 KHNADVVSVSSTGFTGTGLVGEKQWQAIRALRKAGIPMVVATGNYATSSASSSMDLVANNH 397
 DB 301 KHNADVVSVSSTGFTGTGLVGEKQWQAIRALRKAGIPMVVATGNYATSSASSSMDLVANNH 360
 QY 398 LKMTDGTGVTAAHEDAIIVASAKNQVEPDKNIGESKYNIGAFPFKSKITTNEO 457
 DB 361 LKMTDGTGVTAAHEDAIIVASAKNQVEPDKNIGESKYNIGAFPFKSKITTNEO 420
 QY 458 GTKAPSKLKFVYIGKQDQDILGLDRGKIIVMBRIYKDKNAFKKAMDGARAIVNVN 517
 DB 421 GTKAPSKLKFVYIGKQDQDILGLDRGKIIVMBRIYKDKNAFKKAMDGARAIVNVN 480
 QY 518 TVNYNNDNMTLEPAMGYEADGTSQVFSISGDDGVKLMNMINDPKKTTEVRNNKEDFK 577
 DB 481 TVNYNNDNMTLEPAMGYEADGTSQVFSISGDDGVKLMNMINDPKKTTEVRNNKEDFK 540
 QY 578 DKLEQYPIIDMESFNNSKNPNVGEKEIDFKAPPTDKLYKEDIIVPAGSISMGPRIDL 637
 DB 541 DKLEQYPIIDMESFNNSKNPNVGEKEIDFKAPPTDKLYKEDIIVPAGSISMGPRIDL 600
 QY 638 LKPDVSAFGKNIKSTLVINGKSTYGY 664

DB 601 LKPDVSAFGKNIKSTLVINGKSTYGY 627
 RESULT 6
 ID AAU05396
 XX AAU05396 standard; Protein; 26926 AA.
 XX
 AC AAU05396;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human titin (connectin) protein sequence.
 XX
 KW Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
 KW titin-related disease; zebrafish; heart failure; heart disease.
 OS Homo sapiens.
 PN WO200151666-A1.
 XX
 PD 19-JUL-2001.
 XX
 PF 12-JAN-2001; 2001WO-US01212.
 XX
 PR 12-JAN-2000; 2000US-0175787.
 XX
 PA (GENO) GEN HOSPITAL CORP.
 XX
 PI Fishman MC;
 XX
 DR MPI: 2001-451869/48.
 DR N-PSDB; AAS05390.
 XX
 PT Determining if a subject has or is at risk of developing a
 PT titin-related disease or condition, particularly heart failures,
 PT comprises detecting the presence of a mutation in the titin gene -
 XX
 PS Disclosure; Page 57-111; 114pp; English.
 XX
 CC The present sequence representing human titin (also known as connectin)
 CC is described in an invention relating to a novel method for determining
 CC whether a subject has or is at risk of developing a titin-related
 CC disease or condition. The method comprises analyzing a nucleic acid of
 CC a sample from the subject and detecting the presence of a mutation
 CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
 CC titin gene, which indicates that the subject has or is at risk of
 CC developing a titin-related disease. The zebrafish which has a phenotype
 CC similar to mammalian heart failure is used as a model. The method is
 CC useful for detecting an increased likelihood of heart disease, such as
 CC heart failure, in a patient, so that appropriate intervention can be
 CC instituted before any symptoms occur. The method may also be used to
 CC facilitate determination of etiology of an existing heart condition,
 CC such as heart failure, to identify compounds that can be used to treat
 CC or prevent heart conditions, in prenatal genetic screening, e.g. to
 CC identify parents who may be carriers of a recessive titin mutation.
 CC Compounds identified using the methods may be used to treat patients
 CC that have or are at risk of developing heart disease, e.g. heart
 CC failure.
 CC
 XX
 SQ Sequence 26926 AA;
 Query Match 70.8%; Score 3114.7; DB 22; Length 26926;
 Best Local Similarity 9.5%; Pred. No. 4.3e-41;
 Matches 583; Conservative 69; Mismatches 10; Indels 5463; Gaps 479;
 QY 1 Y-----TS-S-----S---E-D-----ALN---ISD-----KEX-- 22
 DB 10392 YCVVNSTGSRKQFCQVNVVDRBGPVGVSDDEVTKDVMVSMKPPLDGSGKITNYI 10451
 QY 8 -----TS-S-----S---E-D-----ALN---ISD-----KEX-- 22
 DB 10452 IEKKEVGKDVWMPVTSASAKTTCKVSKLLEGCKDYIFRIHAEINLYGISDPLVSDSMKAKDR 10511

QY	23	--V--A-----E-----NK-----EK-----H--	30	QY	130	K-----V-----L-Y-----	133
Db	10512	FRVPDAPDQPIVTEVTKDSALVTWKNPHDGGKPIITNYILEKRETMGSKRWARTKDIHPY	10571	Db	11592	KTPILAINPIDRPGEPENLHADKGTFFYLKWRRPDYDGGSPNLHYHVERLKGSDWE	11651
QY	31	-----EN-----I-----	33	QY	134	-----T-Y--DR-----	137
Db	10572	TKFRVPDLLEGCOYEFVRSAENEIGIGDPSPSKVPFAKDPIAKPSPVNPPEAIDTTCNS	10631	Db	11652	RVHGSIKETHYVMDRCVENQIYEFVRQTKNEGGSMDVKTEVVVVKEDIQKPVLDLKLS	11711
QY	34	-----H--S-----A-M-----ET-----	39	QY	138	-----I-----	138
Db	10632	VDLTWQPPRHGGSKILGIVVEYQKVGEWRRAHPTPESCPETKYKVTGLRDGQTYKFR	10691	Db	11712	GVLTVKAGDTIRLEAGVRGKPPPEVAWTKDADTLTRSPRVKIDTRADSSKSLTKAKR	11771
QY	40	-----S-----Q-----D-----F--	43	QY	139	-----F-----N-----G--	141
Db	10692	VLAUNAAGSDPAHVPEPVLVKDRLEPPELILDANMARQCHIKVGDTLRSALIIKGVPPP	10751	Db	11772	SDGGKYVVTATNTAGSFVAYATVNVLDKPGPVNRNLUKIVDVSSDRCTVCWDPDPDGGCEI	11831
QY	44	-----KE-----K-----T-----A--V-IK-----	52	QY	142	-----SA-----IE-----T-----	146
Db	10752	KVTWKEDRDAPTAKRIDVTPVGSKLEIRNAHAHEDGGIYSLTVENPAGSKTVSVKVLVLD	10811	Db	11832	QNYILEKETKEMVWSTYSATVLTPTGTTVTRLIEGNEYIFRVAENKIGTGPTTESKPVI	11891
QY	53	-----E-----KE-----VV-----SK-----	59	QY	147	-----PD-----N--LD-----K--	153
Db	10812	KPGPPDLEVSEIRKDCSYLTWKEPLDDGGSVITNVVVERRDVASAOWSPLSATSKKSH	10871	Db	11892	AKTKYDKPGRDPPPEVTKVSKEEMTVWVNPPEYDGGKSITGYFLEKKEKHSRWVPVNS	11951
QY	60	-----N-----P-V-----I-----D-----	64	QY	154	I-----K-----Q-I--E--GI-----S-----	161
Db	10872	FAKHLNEGNYLFRVAENQYGRGPEVETPKPIKALDPLHPGPPKDLHHVDVDTKEVSL	10931	Db	11952	AIPERMKVQNLLPDHEYQFRVKAENEIGIEPSLPSRPVVAKODIEPPGPTNFRVDT	12011
QY	65	-----N-----N-----T-----S-----N--	69	QY	162	-----S-----VE-----R--AQ--K--	168
Db	10932	VNKPDRDGGSPITGYLVYQEGTQDWIKFKVTNTLECVVTGLQOGKTYRFRVAENIV	10991	Db	12012	TGHSITLGMGKPVYDGGAPIIGYVVMRPKIADASPDEGKFCNAQAOLVRKEFTVTSLD	12071
QY	70	-----E-----E-----AK-----	73	QY	169	-----V--Q-----P-----M-----	172
Db	10992	GLGLPDTTPIECQEKLVPPSVELDVKLI EGLVVKAGTVTRPPAIIIRGVVPVTAKTWTDG	11051	Db	12072	ENQYEFVRCAONQVIGIRPAELKEAIKPKETLEPPEIDLDASMRKLVIVRAGCPIRLFA	12131
QY	74	-----IK-E--E-----N-----SN--K-----	81	QY	173	-----M-----	173
Db	11052	SEIKTDEHYTVETDNFSSVLTIKNCLRRDTGEYQITVSNAGSKTVAVHLTVLVDVGPPT	11111	Db	12132	IVRGPAPKVTRWKVGIQNVVRKGOVDLVDTMFLVIPNSTRDDSGKYSLTUVNPAGEKA	12191
QY	82	-----S-Q--G-----D-----	85	QY	174	-----N-----H-----A-RK-----	178
Db	11112	GPINILDVTPHMTISWQPKDDGSPVINYIVEKQDTRKDTGWVSSGSSKTKLIPHL	11171	Db	12192	VPVNVRLDTPGPVSDLKVDVTKTSCHVSWAPPENDGGSQVTHYIVEKREADRKTWSTV	12251
QY	86	-----YT-----DS-----F-----	90	QY	179	-----EI-----GV-----E-----EA--	185
Db	11172	QXGEYVFRVAENKIGVGPLDSTPTVAKHKFSPSPGPKPVVTDITENAAATVSWTLPK	11231	Db	12252	TPEVKKTSFHTNLVPGNEYFRVTAVNEYGPGVPTDVPKVLASDPLSEDPDRKLEAT	12311
QY	91	-----VNK-----NT-----E-----	96	QY	186	-----ID-----Y-----LK--	190
Db	11232	SDGSPITGYMERREVTGKVRVNRKTPITADLKFRVTGLYEGNTYEFVRFAENLAGLSKP	11291	Db	12312	EMTKNSATLAWLPPLRDGAKIDGVIISYREBEQPADRWTEYSVVVKDLSLVVTGLKEGKK	12371
QY	97	-----NPK--K-E--D--K-----VV-----	105	QY	191	-----SI-----	190
Db	11292	SPSSDPIKACRPIKPPGPINPALKDKSRETADLWTKPLSDGGSPILGYYVECOQPGTA	11351	Db	12372	YKFRVAARNAVGVSUPREAGVYEAKQLPPKILMPQITIKAGKCLRIBAHVYKPKHP	12431
QY	106	-Y-----I--A-----E--F-----KD-----	112	QY	191	-----	192
Db	11352	QMNINKDELIRQCAFVRPGLIEGNEYFRFRIKAANIVGEGEPRELAESVIAKDILHPPEV	11411	Db	12432	TCKWKKEDEVVTSHLAVHKADSSSILIIKDVTBKDSGYSLTAENSSGTDQIKIKVVV	12491
QY	113	-----K--E--S--G-----EK-----	118	QY	193	-----	192
Db	11412	ELDVTCRDVITVRVGTIRILARVGRPEPDITWTKEGKVLVREKREVDLIQDLPRVELOI	11471	Db	12492	MDAPGPPPPDISIDADACSLSWHIIPELDDGGSNITVIVEKCDVSRGMDWTALASVTK	12551
QY	119	-----AI-----KE-----L--	123	QY	193	-----N-----A--PFG--KN-----	199
Db	11472	KEAVRADHKYIISAKNSGHAQGSALVNVLDKPGPCQNLKVTNVTKENTISWENPLDN	11531	Db	12552	TSCRVGKLIPGOEYIFRVAENRFGISBPLTSPKVAQOPFGVGPSEPKNARVTKVNDKCI	12611
QY	124	-----S--S-----L-K-----N-----T	129	QY	200	F-----DG-----R-----GMV-----I-S	208
Db	11532	GGSEITNFIVEYRKNQKGSIVASDVTKRLIKANLLANNEYFRVCAENKVGVGPTIET	11591	Db	12612	FVAMDRPDSGGSPIIGYLIERKERNLLWVKANDTLVRSTEYPCAGLEGYEFRIYA	12671
				QY	209	-N-----ID-----T-----GT-----	214

```

Db 12672 LNKAGSPPSKPTXYVTARMPVDPGKPEVIDVTKSTVLSLIMARPKDGGSKIIGYFVEA 12731
Qy 215 -D-Y-R-----H-----KA-----M- 221
Db 12732 CKLBGDKKVRNCNTAHPQIPOEYATATGLEEKAQYQFRAIARTAVNISPPSPSDPVTILA 12791
Qy 222 -RID-----D-----D-----A-----KASW-R 232
Db 12792 ENVEPRIDLSVAMKSLTLVAKGTNVCDAITVFGKMPVSMKKDGTLLKPEGIKMMQOR 12851
Qy 233 -F--KKE-DL--K--G--T-----DK-----N--Y----- 245
Db 12852 NLCTLELFSVNRKDSGYTITAENSSSGSKATIKVLKDKGPASVXINMGYGBRAWLS 12911
Qy 246 W--L--S-----DK-----IP-----H--AF-----N-Y 257
Db 12912 WEPLEDGGSEITNYIVDKRETSRPNMAQVSATVPITSVCEKULEGHEYOFRICAENKY 12971
Qy 258 -----Y-----N-----GG-KIT----- 264
Db 12972 GVGDPVTEPAIAKPIPDPCRCPPVLSNITKDHMTVSMKPPADDGSSPITGYLEKRE 13031
Qy 265 -V-----E----- 266
Db 13032 TQAVNMTKVNKKPIIERTLKATGLOEGTEYEFRTAINKAGPGKPSDASKAAYARDPQYR 13091
Qy 267 -K-YD-----DG-----R 272
Db 13092 PAPAPFKVDYDTRSSVSLSGKPAYDGGSPIIIGLYVEKRADSDNWRNCLPNOLQKTR 13151
Qy 273 -D-Y-F-----DP-----H-----GMH-----I--AG 284
Db 13152 FEVTGLMEDTOYQFRVAVANKIGYSDPSDVPDKHYPPKDILIPRGEHDADLRKTLILRAG 13211
Qy 285 -----I-----L-----AG----- 288
Db 13212 VTMRLVYPVKGRPPPKITWSKPNVNLRIGLDKSTDFOTFLCENNVKRYDAKYLITL 13271
Qy 289 -N-----DT-----E-----Q--D- 294
Db 13272 ENSCKKEKTYIVKVLDPGPPINVTAKESKDSAYVTWEPPIIDGSGPIINYVQKRDA 13331
Qy 295 -----I-----K-----N-----F--N--GI--D--G-----I--A----- 305
Db 13332 ERKSWSTVTTECSKTSFRVPNLEBCKSYFFRVFAENEXIGDPGEBTRDAVASQTPGEPVV 13391
Qy 306 -----P-----N-----A--Q----- 309
Db 13392 DLKRVSVSKSCSIGMKKPHDGGRIIGYVVDPLTEENKMQRVMSKLSLOYSAKDLTEG 13451
Qy 310 -----I-----F----- 311
Db 13452 KEYTFRVAENENGEPTSEITTVARDDVAPDLKGLPDLCYLAKENSNFLKIPKIG 13511
Qy 312 -SYKM-----Y-----SDAG----- 320
Db 13512 KPAPSVMKGEDPLADTRVSESSAVNTLLIYDCKQSDAGKYTTLLKXNAVGTKEGTI 13571
Qy 321 S-----G--F-----A-----G-----D----- 326
Db 13572 SIKVVGKPGIPTGPIKPEDEVTAEMTLKMAPPKDGGSEITNYILEKSDVNNKMYTCAS 13631
Qy 327 -ETMF-----H-----A-----I--E--DS--I--KH-----N--VDV- 343
Db 13632 AVQKTTFRVTRLHGEKMEYTFRVSABNKYVGEGKSEBIVARHPFDVADAPPNNIVDR 13691
Qy 344 -VSV-----S--SG-----F--TG--T--GL----- 355
Db 13692 HDVSVLWTDBKKTGTGSGPTIYHLKERNKSLMLWKRANKPTDIRMDPFVTLTGLELEYEF 13751
Qy 356 -VG-----E--KY----- 360

```

```

Db 13752 RVMAINLAGVKSLSLPESEPVVALDIPDPGKPEVINITRNSVTLIMTEPKYDGHKLTCY 13811
Qy 361 -W-----Q--A-----I-----PALR-K--AG-I--P-----MV 375
Db 13812 IVERKDLPSKSMKANHNVPCEAFVTVDLVEGGKYEF--IRAKNTAGAISAPSESTETI 13870
Qy 376 V-----A-----T--G-----N----- 380
Db 13871 ICKREYAPPTIVDPTIKDGLTIKAGDTYVUNAISILGKPLPKSSWSKAGKDIRPSDITO 13930
Qy 381 -YAT----- 385
Db 13931 ITSTPTSMLTIKYATRCKDAGEYTTATNPFGTKEHVKTVDLPDPGPPVEISNVAE 13990
Qy 386 -S--S-----SS--W-----DL--VA-----N----- 395
Db 13991 KATLTWTPLEDGSPKSYILEKRETSRLMTVISEDIOSCRHVAATKLIOGNEYIFRVS 14050
Qy 396 -NHL-----WTD--TG-----NVR--TAA-----H 412
Db 14051 AVNHYGGEVQSEBPVAMVDRFPGPPEKPEVSNVTKTATVSMKRPVDDGSEITGYH 14110
Qy 413 -E-----D-----A--I-----A--V----- 418
Db 14111 VERREKSLRWVRAIKTPVSDLRCKVTGLOGSTYEFRVSANRAGICGPSEASDSVLMK 14170
Qy 419 -A-----SA-----K-----N-----Q-----T----- 425
Db 14171 DAAYPGPSPNPVHTDTTKKSASLAWGKPHYDGLIETGYVVEHQVDEAMIKDTTGTA 14230
Qy 426 -V-EF--D--K-----VN-----I----- 433
Db 14231 LRITQFVPLDQTEKYNFRISAINDAGCEPAVIVPELVERMAPDFELDAELRRTLV 14290
Qy 434 -G-----G-----ESF-----KY----- 440
Db 14291 VRAGLSIRIFVPIKGRAPAVETWKONINLKNRANIENTESFTLLIPECGRVOTGKFMV 14350
Qy 441 -----R-----N-----I--G----- 444
Db 14351 TIENPAGKSGPVNVRLDTPGVPVNLRPDTDKDSVTLHMDDLPLIDGSRITNYIVEKR 14410
Qy 445 -----A-----FF----- 447
Db 14411 EATRKSYSTATTCKHKCTYATVGLSEGCCEYFFRVMAENEYIGEPPTTTEPVAKEAESP 14470
Qy 448 -----D-----K--SK-----ITT----- 454
Db 14471 PDSLINIMDITKSTVSLAMPKPKDGGSKITGYVIEAQRKSGSDOWHTITTYKGLCEVVRNL 14530
Qy 455 -----N-----E--D--G-----TKA----- 461
Db 14531 TEGEYTFQVMAVANSAGRSAPRESRPVIVKEQTLPELDLRIYQKLVIAKAGDNIIVEI 14590
Qy 462 -PS--K--LK-----F-----Y-----I--GK- 472
Db 14591 PVLGRPKPTVWKKGDOILKQTORVNFETTATSTILINECVNSDSGPPPLTANRIVGEV 14650
Qy 473 G--Q-----D-----Q--D----- 477
Db 14651 GDVITIOVNDIPGPPIPIKPEVSSDPVTFMSMDPENBDGVPISNVVEMRQJDSITWW 14710
Qy 478 -L--I-----GL----- 481
Db 14711 ELATTVIRTTYKATRLTTGLEYOFRVAKQNRGVGPGITSAMIVANYPFKVPGBPPTQV 14770
Qy 482 -----D--L-----R-----G--K-----I 487
Db 14771 TAVTKDSMTISMEPLSDGSPILCYHVERKERNGIIMQTVSKALVPDGNIFKSSGLTDGI 14830
Qy 488 A--V-----M-----D-----R--I-----YT----- 495
Db 14831 AYEFRVIAENMAGSKSKSPSEPLALDIPDPGKVPVPLNITRHTTVTLKNAKPEYTOGFK 14890

```

QY	496	-----K-DL-----	-----KNA-----	501	QY	601	-----EK-----E-----ID--F-----KP-----A-----P--	610
Db	14891	ITSYIVKRDLPNGRWLKNFNSILENEFTVSGLTEDAAVEFRVIAKNAAGAISPPSEPS	14950	Db	15971	SKITGYIVEKELPEGRWKASFTNIIDTHFEVTGLVEDHRYEFRVIAAAGVFPSE	16030	
QY	502	-----FK-----KA-----M-----	506	QY	611	-----DT-----D-----K-----EL-----	616	
Db	14951	DAITCRDVEAPKIKVDVKFDTVILKAGEAFRLADVSGRPPPTMEWSKDGKELEGTA	15010	Db	16031	STGALTARDEVDPPRISMDPKYKDTIVHAGESFKVDADIYKGPPTIQWIKDQELSNT	16090	
QY	507	-----D-----	507	QY	617	-----Y-----K-----E-----D-----II-----	622	
Db	15011	LEIKIADFSTNLVKNKSTRRDSGAYTLTATNPGGFAKHI FNVKVLDRPPEGGLAVTEV	15070	Db	16091	ARLEIKSTDFATSLSVKDAVRVDSGNYILKAKNVAGERSVTNVVKVLDPRPPEGPVVIS	16150	
QY	508	-----K-----GA-----R-----A-----I-----	513	QY	623	V-----P-----AGS-----TS-----W-----	630	
Db	15071	TSEKCVLSWFPLDGGAKIDHYIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNIEYIF	15130	Db	16151	GVTAEKCTLAWKPPLODGGSDIINYIVERRETSLRWTVVDANVOTLSCKVTKLLEGNEY	16210	
QY	514	-----M-----V-----VN-----T-----	519	QY	631	-----G-----	632	
Db	15131	RVMAVNYGVEPLESEPVLANPYGPPDPKPNPEVTTITKDSMVVCGHDPDSDGSEII	15190	Db	16211	TFRIMAVNKGVEPLESEPVVAKNPFVVVPDAPKAPVTTVTTKDSMIVVWERPASDGGSE	16270	
QY	520	NY-----	522	QY	633	-----R-----I-----DL-----L-----	637	
Db	15191	NYIIVERDKAGORWIKCNKKTLDLRYKVSGLTEGHEYFRIMAENAAGISAPSPSPFY	15250	Db	16271	ILGYVLEKRDKEGIRWTRCHKRLIGELRLRVGTGLIENHDYEFVRVSAENAAAGLSEPPSPA	16330	
QY	523	-----N-R-D-----N-----W-----	527	QY	638	-----LKP-----D-----V-----SAP-----G-----K-----N-----	648	
Db	15251	KACDTVPKPGPNRVLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEDEMQIVTP	15310	Db	16331	YQACDPIYKPGPNPNKVIDITRSVFLWSKPIYDGGCEIOGYIIVEKCDNVNGEWTMC	16390	
QY	528	-----TE-----L-PA-----M-----	533	QY	649	-----I-KST-----L-----N-----	654	
Db	15311	PAGLKATSYTITGLTENQEKIRIYAMNSEGLGEPAVPGTPKABDRMLPPEIELDADLR	15370	Db	16391	TPPTGINK-TNIEVEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKEAPDIDL	16449	
QY	534	-----	534	QY	655	-----VIN-----GK-----S-T-----Y-----	662	
Db	15371	KVVTIRACCTLRFVPIKGRDPDVKWARDHGESLDKASIESASYSYTLIIIGNVNRFDG	15430	Db	16450	ELRKIINIRAGGSLRFLFPIKGRPTPEVKWGVGDGEIRDAIIDVTSFTSLVLDNVNRY	16509	
QY	535	Y-----E-----A-----D-----E-----G-----	540	QY	663	--G-Y 664		
Db	15431	KYILTVENSGSGSAFNVNRVLDTPGPPQDLKVAKEVTKTSVTLTWDPPLDGGSKIRNYI	15490	Db	16510	DSGKY 16514		
QY	541	-----T-----K-S-----Q-----VF-----SI-S-----	549					
Db	15491	VEKRESTRKAYSTVATNCHTKTSWKVDQLQEGCSYFRVLAENEYIGIUPAETAESVKASE	15550					
QY	550	-----G-----D-----DG-----VKL-----W-----	557					
Db	15551	RPLPPGKITLMDVTRNSVSLSWKEPHEHGGSRILGYIVEMQTKGSDKWATCATVKVTEAT	15610					
QY	558	-----N-----M-I-----N-----	561					
Db	15611	ITGLIOGEYSFRVSAQNEKISDPRLSPVPIANDLVIPPAFKLLFNTFTVLAGEDLKV	15670					
QY	562	-----P-----D-----K-----K-TE-----VK-----R-----N-----	571					
Db	15671	DVPPIGRPTPAVTHKONVPLKQTRVNAESTENNSLLTIKDACREDVGHVVKLTNSAG	15730					
QY	572	-----N-----K-E-----D-----F-----K-D-----	578					
Db	15731	EAIETLNVILDKPGPPTPGVKMDVTDADTSITLSWGPVKYDGGSSINNYIIVEKRDSTTT	15790					
QY	579	-----K-----L-----E-Q-----Y-----Y-----	585					
Db	15791	WQIVSATVARTTIKACRLKTGCEYQFRIAENRYGKSYTLNSEPVAQYPPKVPGPPTG	15850					
QY	586	I-----D-ME-----S-----P-----NS-----NK-----P-----	596					
Db	15861	VVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERNISILWKLKNTPIPTQTKFTTGLEE	15910					
QY	597	-----N-VG-----	600					
Db	15911	GVEYEFVSAENIVGIGKPSKVSCEYVARDPCDPPRPEAIIVTRNSVTLQWKPTYDGG	15970					

QY	601	-----EK-----E-----ID--F-----KP-----A-----P--	610
Db	15971	SKITGYIVEKELPEGRWKASFTNIIDTHFEVTGLVEDHRYEFRVIAAAGVFPSE	16030
QY	611	-----DT-----D-----K-----EL-----	616
Db	16031	STGALTARDEVDPPRISMDPKYKDTIVHAGESFKVDADIYKGPPTIQWIKDQELSNT	16090
QY	617	-----Y-----K-----E-----D-----II-----	622
Db	16091	ARLEIKSTDFATSLSVKDAVRVDSGNYILKAKNVAGERSVTNVVKVLDPRPPEGPVVIS	16150
QY	623	V-----P-----AGS-----TS-----W-----	630
Db	16151	GVTAEKCTLAWKPPLODGGSDIINYIVERRETSLRWTVVDANVOTLSCKVTKLLEGNEY	16210
QY	631	-----G-----	632
Db	16211	TFRIMAVNKGVEPLESEPVVAKNPFVVVPDAPKAPVTTVTTKDSMIVVWERPASDGGSE	16270
QY	633	-----R-----I-----DL-----L-----	637
Db	16271	ILGYVLEKRDKEGIRWTRCHKRLIGELRLRVGTGLIENHDYEFVRVSAENAAAGLSEPPSPA	16330
QY	638	-----LKP-----D-----V-----SAP-----G-----K-----N-----	648
Db	16331	YQACDPIYKPGPNPNKVIDITRSVFLWSKPIYDGGCEIOGYIIVEKCDNVNGEWTMC	16390
QY	649	-----I-KST-----L-----N-----	654
Db	16391	TPPTGINK-TNIEVEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKEAPDIDL	16449
QY	655	-----VIN-----GK-----S-T-----Y-----	662
Db	16450	ELRKIINIRAGGSLRFLFPIKGRPTPEVKWGVGDGEIRDAIIDVTSFTSLVLDNVNRY	16509
QY	663	--G-Y 664	
Db	16510	DSGKY 16514	
	RESULT 7		
	ABG74786		
ID	ABG74786	standard; Protein; 31267 AA.	
XX			
AC	ABG74786;		
XX			
DT	05-JUN-2003	(first entry)	
XX			
DE	Human RGS11 protein.		
XX			
KW	RGS11; human; screening; cardiant; antianginal; gene therapy;		
KW	heart disorder; cardiac ischaemia; heart failure; angina.		
OS	Homo sapiens.		
XX			
PN	WO2002103355-A1.		
XX			
PD	27-DEC-2002.		
XX			
PF	17-JUN-2002; 2002WO-JP06019.		
XX			
PR	18-JUN-2001; 2001JP-01803038.		
XX			
PA	(TAKE) TAKEDA CHEM IND LTD.		
XX			
PI	Koyama N, Tanida S, Yamamoto K;		
XX			
DR	WPI; 2003-167557/16.		
DR	N-FSDB; ABX13540.		
XX			
PT	Screening compounds regulating RGS11 expression and activity for		
PT	prevention and treatment of heart disease		

RESULT 7

ABG74786

ID ABG74786 standard; Protein; 31267 AA.

XX ABG74786;

AC 05-JUN-2003 (first entry)

DT DT

XX Human RGS11 protein.

DE RGS11; human; screening; cardiant; antianginal; gene therapy;

XX heart disorder; cardiac ischaemia; heart failure; angina.

XX Homo sapiens.

OS WO2002103355-A1.

XX 27-DEC-2002.

XX 17-JUN-2002; 2002WO-JP06019.

XX 18-JUN-2001; 2001JP-0183038.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Koyama N, Tanida S, Yamamoto K;

XX WPI; 2003-167557/16.

XX N-PSDB; ABX13540.

XX Screening compounds regulating RGS11 expression and activity for

PT prevention and treatment of heart disease

XX Claim 1; Page 59-261; 321pp; Japanese.

CC This invention describes a novel method for screening compounds for their
 CC ability to regulate the activity and expression of human RGS11 and its
 CC partial peptides and salts, by observing the expression or activity of
 CC RGS11 in the presence or absence of the test compound. The products of
 CC the invention have cardiant and antilanginal activity and can be used for
 CC gene therapy. The methods and compositions are useful in the prevention,
 CC treatment and diagnosis of heart disorders such as cardiac ischaemia,
 CC heart failure and angina. This sequence represents the human RGS11
 CC protein described in the disclosure of the invention.

XX Sequence 31267 AA;

Query Match 70.6%; Score 3104.2; DB 24; Length 31267;

Best Local Similarity 9.4%; Pred.No.1,6e-40; Indels 5628; Gaps 490;

Matches 594; Conservative 59; Mismatches 11; Indels 5628; Gaps 490;

QY 1 Y-----PV-----V-----LAD-----7
 Db 15617 YCVVENVSTGRKGFCCQVNVVDHGPVGVSPFDEVTQDYNVISMKPELDGGSKITNYI 15676
 QY 8 -----TS-S-----S--E--D-----ALN--ISD-----KEK--22
 Db 15677 IEKKEVGKDVMPVTSASAKTTCKVSKLEGKDYIFRIHAENLVGISDPLVSDSMKADR 15736
 QY 23 --V--A-----E-----NK-----EK-----H--30
 Db 15737 FRVDPADQPIVTEVTQDSALVTWKKPHDGGKPIITNYILEKRETMKRWMAVTKDPIHPY 15796
 QY 31 -----EN-----I-----33
 Db 15797 TKFRVPDLLECCQYEFVSAENIGIDPSPSPKVFAPKADPIAKSPVFNALDITGNS 15856
 QY 34 -----H--S-----A-M-----ET-----39
 Db 15857 VDLTPQPRHDGSKILGIVIEYQKVGDEWRBRANHTPESCFETKYKVTGLRDGQYKFR 15916
 QY 40 -----S-----Q-----D-----F--43
 Db 15917 VLAANAAGESDPAHVPEVLVKDLPELILDANMARQHIKVGDTLRLSAIKGVFP 15976
 QY 44 -----KE-----K-----T-----A--V--IK-----52
 Db 15977 KVTWKEDRDAPTKARIDVTGVGSKLEIRMAHEDGIGYSLTVENPAGSKTVSVKVLVD 16036
 QY 53 -----E-----KE-----VV-----SK-----59
 Db 16037 KRGPRLDVESEIRKDCYLTKWKEPLDGGSVITNYYERRDVASQWSPISATSKKSH 16096
 QY 60 -----N-----P--V-----I-----D-----64
 Db 16097 FAKHLNEGNOYLFRVAENQYGRGPFVETPRKIKALDPLHPGPKDLHADVKTESVL 16156
 QY 65 --N-----N-----T-----S-----N--69
 Db 16157 VMNRPDRDGGSPITGYLYEYQEGTQDWIKKFTVTNLECVVTGLQGGKTYFRVKAENIV 16216
 QY 70 -----E-----E-----AK-----73
 Db 16217 GLGLPDTTPIIECEKLVPSVELDVKLEGLVVAAGTTVPFPAIRGVPTAKWTTDG 16276
 QY 74 --IK--E--E-----N-----SN--K-----81
 Db 16277 SEIKTDEHYTETDNFSSVLTIKNCRLRDTGEGYQITVSNAGSKVAHVLTLVDPGPPT 16336
 QY 82 -----S-Q-----G-----D-----85
 Db 16337 GPINILDVTPPEHMTISWQPPKDDGSPVINYIVKQDTRKDTWGVSSGSSKTKLKIPHL 16396
 QY 86 --YT-----DS-----F-----90

Db 16397 QKCEYVFRVAENKIGVPELDSTPTVAKHKSPSPSPCKPVVTDITENAAVSWTLPK 16456
 QY 91 -----VNK-----NT-----E-----96
 Db 16457 SDGSPITGYMERREVTGKMKVRVNTPIADLKFRVGLYEGNITYFRVFNLAGLSKP 16516
 QY 97 -----NPK--K--E--D--K-----VV-----105
 Db 16517 SPSSDPIKACRPKIPGPPINPKLKDSDRETLADLWMTPLSDGSPILGVVVECCQKPTA 16576
 QY 106 -Y-----I--A-----E--F-----KD-----112
 Db 16577 QMNRINKDELIRQCAFVPGLEIGNEEYFRKIANIVGEGSPRELAESVIKDTLHPPEV 16636
 QY 113 -----K--E-----S--G-----EK-----118
 Db 16637 ELVDTCSDVITVRVGQITRILARVKGPEPDITWTKGKVLVRKRVLDIODLPRVELQI 16696
 QY 119 -----AI-----KE-----L--123
 Db 16697 KEAVRADHGKYLISAKNSSGHAQGSALVNVLDNRGPGCNLKVNTVTKENTISMENPLDN 16756
 QY 124 -----S--S-----L--K-----N-----T--129
 Db 16757 GGEITNFIYEVRKPKNGKWSIVASDVTKRLIKANLLANNEYFRVCAENKVGVPITET 16816
 QY 130 K-----V-----L--Y-----133
 Db 16817 KTPILAINPIDRGPENLHADKGTFFVYKWRPRPDYDGSPLSYHVERRLKGSDDWE 16876
 QY 134 -----T--Y--DR-----I--F-----N--G-----S--142
 Db 16877 RVHKGSIKETHYMDRCVENQIYEFRVQTKNEGSESQWVTEBVVVKEDLQKPVLDLKL 16936
 QY 143 -----A--I-----E-----T--T--P-----148
 Db 16937 GVLTVKAGDITRLEAGVGRPFPEVAMTKODATDLTRSPVKIDTRADSSKPSLTKAKR 16996
 QY 149 -D-----N--LD-----KI-----K--155
 Db 16997 SDGKYVVTATNTAGSFVAAYATVAVLDPGQVRNLKIVDSSDRCTVCMPPEDDGCET 17056
 QY 156 Q-----Q-----IEG-----I--160
 Db 17057 QNTILEKCETKRMWSTYSATVLTGTTVTRILIGNENYIRVRAENKIGTGPTESKPVI 17116
 QY 161 -----S-----SV-----E--R-----165
 Db 17117 AKTKYDKRPRDPPEVTKVSKEBMTVMNPREYDGSINGYFLEKKEKISTRWVPVKS 17176
 QY 166 A-----Q--KVQ-----P-----171
 Db 17177 AIPERRMKVONLPRDHEYOFKVAENEGIGEPSLPSRPVAVKADPIEPGPPTFRVVD 17236
 QY 172 -----M-----N-----H-----A--RK-----172
 Db 17237 TKHSITLGMGKPVYDGGAPIIIGYVEMRPKIADASPDEGMRKCAAAQOLVRKEFTVSLD 17296
 QY 173 -----M-----N-----H-----A--RK-----178
 Db 17297 ENGEYFRFMSLKHGHVSWAPRENDGSGVTHYIVBKRBADRKWSTVTPPEVKTSFHV 17356
 QY 179 -----I-----E-----179
 Db 17357 NLVPGNEYFRVAVNEYGPGVPTDVPKPLASDPLSEPPPKLEVTENTKONSATLAML 17416
 QY 180 -----I-----E-----180
 Db 17417 PPLBDGAKKIDGYITTSREBQPADRWTEYSVVKDLSLVTTGLKEGKKYKFRVAARNAV 17476
 QY 181 -----GV-----E-----EA-----185
 Db 17477 VSLPRAEGVYEAKEQLLPKILMPQIITIKAGKULRIEAHVVGKPHPTCKMKKGDEBIV 17536

QY	186	-----	185	QY	307	-----NA-----Q-----	309
Db	17537	TSSHLAVHKADSSILIIKDVTRKSGYYYSLSLTAENSSGTDQKIKVVMNDAPGPPQPPD	17596	Db	18617	DPLATDTRVSVSSAVNTLIVYDCKQSDACKYITILKNVAGTKEGTISIKVVGKPGIPT	18676
QY	186	ID-----Y-----L-----K-S-----	191	QY	310	I-F-----S-----Y-----K-----	314
Db	17597	ISDIDADACSLSWHIPLEGGSNITNYIIVEKCDVSRGDWVTALASVTKTSCRVGKLIPOQ	17656	Db	18677	GPIKFDEVTAEAMTLKWAPKDDGSEITNYLEKRDVNNKNWVTCASAVQKTTFRVTRL	18736
QY	192	I-----N-----A-PFG-----KN-----F-----DG-----	202	QY	315	M-----Y-----S-----DA-----	319
Db	17657	EVIFRVAENRRGISEPLTSPKNVAQFPGVPSEPKNARVTKVKNKDCIFVAMDRPDSOGG	17716	Db	18737	HEGMEYTRFVSAENKYGVGEGLSKSEPIVARHPFDVPDAPPPENIVDVHRHDSVSLTWDPK	18796
QY	203	-----R-----GMV-----I-S-N-----	209	QY	320	GS-----G-----F-----AG-----	325
Db	17717	SPIIGYLIBERKERNLLWKANDTLVRLTEYPCAGLVEGLEYSFRYIALNKGAGSPPSKP	17776	Db	18797	KTGSGPIYGYHLEFKERNLLWKANKTPIRMDFKVTGLTEGLEBYFRVMAINJAGVK	18856
QY	210	ID-----T-----GT-----D-Y-R-----	217	QY	326	D-----E-----	327
Db	17777	TEYVTARMPDPPGKPEVIDVTKSTVSLIWARPKHDGSKIIGYFVEACKLPDCKWVRN	17836	Db	18857	PSLPSEPVVALDPIEAPTIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGKD	18916
QY	218	H-----KA-----M-----RID-----	224	QY	328	T-----M-----F-----HA-----	332
Db	17837	TAPHQIQBEYTATGLEEKAQYQFRAIARTAVNISPPSESPDPVTILAENVPPRIDLSVA	17896	Db	18917	IRPSDITQITSTPTSSMLTIKYATRKDAGEVITATNPFGTKVEHVKVTVLDVPVGPV	18976
QY	225	D-----D-----A-----KSM-R-----F-----K-----	234	QY	333	I-----ED-S-IK-----H-----	339
Db	17897	MKSLLTVKAGTNVCLDATVFGKPMPTVSWKBDGTLKPAEGIKMAMQORNLCTLELFSVRN	17956	Db	18977	EISNVSAEKATLTWTPPLEDGGSPIKSYILEKRETSRLLTWTVVSEDIOSCRHVATKLIQ	19036
QY	235	KE-DL-----K-G-----T-----DK-----N-Y-----W-L-----S-----	248	QY	340	N-----N-----VD-----V-----VS-----	345
Db	17957	KOSGDYTTAENSSGKSATIKLVLDKGPASVKINKYIDRAMLSWEPPLEDGSGEI	18016	Db	19037	NEYIFRVAENVHYKGEPVQSEPVQKVDVRFPGPPGPEKPEVSNVTNTATVSKPEPVDG	19096
QY	249	DK-----IP-----H-AF-----N-Y-----	257	QY	346	V-----S-----S-----G-----F-----	350
Db	18017	TNVIDKRETSRPNQAQVATPITSCSVKEKIEGHEYQFRICAENKYGVGDPVFTEPAI	18076	Db	19097	GSEITGYHVERREKSLRWRAIKTPVSDLRCKVTGLQEGSTYEFVSAENRAGIGPPNP	19156
QY	258	Y-----N-----GG-KIT-----V-----	265	QY	351	T-----T-----G-----G-----	357
Db	18077	AKNPYDPPGRCPPVISNTIKOHTMVTWSKPPADGGSPITGYLLEKRETOAVNWKVNRK	18136	Db	19157	PGPPSNPHVDTTKKSASLANGKPHYDGGLEITGYVVEHOKVGDAMWKDTGTALRITQ	19216
QY	266	E-----E-----K-YD-----	269	QY	358	EKY-----	360
Db	18137	PIIERTLKATGLOEGTEYEFRTAINKAGGKPSDASKAAYARDPOYPPGPAFPKYDT	18196	Db	19217	FVVPDLQTKEXNFRISAINDAGVGEPAVDPDVEIVEREMAPDPDELDAELRRLTVRAGL	19276
QY	270	DG-----R-D-----Y-----	274	QY	361	W-----W-----QA-----I-----R-----	365
Db	18197	TRSSVSLSGKPAYDGGSPIIGYLVVEVKRADSDNMVRCNLQNLOKTRPEVTGLMEDTQY	18256	Db	19277	SIRIFVPIKGRPAPEVTWTKONLNKRANIENTESFTLLIIPECNRYDTGKFMVTIENP	19336
QY	275	F-----DP-----H-----G-M-----	280	QY	366	A-----LR-----	368
Db	18257	QPRVAVNKIGYSDPSDVPDKHYPKDILIPPEGELDADLRKTLILRAGVTMLRYVPVKGR	18316	Db	19337	AGKSGFVNVRLDTPGPNLRLPDTITKDSVTLHWDLPIDGGSRITNYIIVEKREATR	19396
QY	281	H-----I-----AG-IL-----	286	QY	369	K-----K-----A-----GI-----P-----	373
Db	18317	PPPKITWSPKNVNLDRIGLDIKSTDFTLRCENVNKYDAGKYILTLNENSGCKEYITIV	18376	Db	19397	SYSTATTKCHKTKYKVTGLSECEYFFRVAENEGIGEPTETTEPVPKASEAPSPDSLN	19456
QY	287	A-----A-----G-----N-----D-----TB-----	292	QY	374	M-----V-----V-----A-----T-----G-----N-----	380
Db	18377	VKVLDTPGPNVTVKEISKDSAYVTWEPPIIDGGSPIINYYVVKQDAERKSWSTVTTEC	18436	Db	19457	IMDITKSTVSLAWPKPKHDGSGKITGYVIEAQKSGDQWHTITTVKGLECVVRNLTEGEE	19516
QY	293	-----Q-----D-----	294	QY	381	Y-----A-TSA-S-----S-----	387
Db	18437	SKTSFRVNLBECKSYFFRVAENEGYIGDPGETRDAVKASQTPGPNVLDKVRSVSKSSC	18496	Db	19517	YTFQMVNSAGRSAPRESRPVIVKEQTMPLPELDLRLGYQLVIAKAGDNKIVEPVLGR	19576
QY	295	I-----K-----N-----F-----	298	QY	388	S-SW-----D-L-----V-----A-----N-----	396
Db	18497	STGWKPKHSDGSGRIIGYVVDLFTENKQVRNKSLSLOYSAKDLTEGKEYTFRVSAENE	18556	Db	19577	PKPTVMKKGDOILKQTORVANFETTATSTILNINECVSRSDSGPYPLTARNIVEGVDVIT	19636
QY	299	NG-----I-----D-G-----I-----AP-----	306	QY	397	H-----H-----LK-----M-----TD-----	402
Db	18557	NGEGTSEITVWARDVWAPDLDLKGLPCLYLAKENSFRLKIPKIGKPAFSPVSKMGE	18616	Db	19637	IOVHDIPGPPTGPIKDFEVSDFVTFSWDPPDGGVPISNVVEMRQDSTTWVELATT	19696
QY				QY	403	-----TG-----N-----V-----T-----	407

Db 19697 VIRTAKATRLTGTLEQFVRVKAQNRVGVGITSACIVANVPFKVPGPQIPQIRPGXR 19756
 Qy 408 -----RT-----A-----A-----H----- 412
 Db 19757 ARIAEFVRVIAENNAAGSKSPKSPBMLALDIPDPGRVPLNTRHTVTLKMAKPEYTG 19816
 Qy 413 -----E--D-----A--I-----AV-----AS-----AKI----- 423
 Db 19817 GFKITSYIVEKRDLPNGRWMLKANSILNENEFVSGLTEDAAIEFVRVIAKQAAGNISPPS 19876
 Qy 424 -----Q-----TV-----E--F-----D----- 429
 Db 19877 EPSDAITCRDDVEAPKIKVDVKFKDVTILKAGBAFRLBADVSGRPPTMEWSKDGKELEG 19936
 Qy 430 -----K-----VN-----I----- 433
 Db 19937 TAKLEIKIADFTSLVNKDSSTRDSGAYTLTATNPGFPAKHIFNVKVLDRPGPEGLAV 19996
 Qy 434 -----GG-----E-----S----- 437
 Db 19997 TEVTSEKCVLSWFPPLDGGAKIDHIYIVQKRETSRLANTVASEVQVTKLVTKLLKNE 20056
 Qy 438 -----KY----- 440
 Db 20057 YIFRMAVNAKYGVEGLESEPVLANPYGPPDPKNPEVTTITKDSMVVCMGHPDSDGS 20116
 Qy 441 -----R--N-----I-----G--A----- 445
 Db 20117 EIINVIERRDKAGORWIKCKNKTLLDLRYKVSGLTEGHEVEFRIMAENAGISAPSPS 20176
 Qy 446 -----F-----D-----K-----S----- 450
 Db 20177 PFYACDVFPRPGPNRVLDTSRSSISIAMNKPIYDGSSEITGYWEIALPEDEKQI 20236
 Qy 451 -----K-----IT--T-----N--E--D-----GT--KA-----PS--KL-- 465
 Db 20237 VTPRAGLAKATSYITLTGLENQYKIRIYANNSGEGBALVGPTRKADRLMPEIELDA 20296
 Qy 466 -----K-----FV-----Y-----I----- 470
 Db 20297 DLKRVVTIRACCTLRLLFVPIKGRPAPEVMAWDRHGESLDAKSIESTSSYLLIGNVNR 20356
 Qy 471 -----GK-----G--D-----QD-----I-----G----- 480
 Db 20357 DSGKYLTVENSSGSKAFVNVRLDTPGPQDLKVEKVTYSVTLTWDPPLDGSKIK 20416
 Qy 481 -----L-----D-----L----- 483
 Db 20417 NYIVEKRESTRKAVSTVATNCHKTSWKVDQLOEGSYFVRVLAENEYIGILPAETAESVK 20476
 Qy 484 -----R--GKIAVMD-----RI--Y----- 494
 Db 20477 ASERPLPFGKITLMDVTRNSVLSWEKPEHDGSRILGIYIVEMQTKSGDKMWCATVAVT 20536
 Qy 495 -----T-----KDL-----K--N-----A-- 501
 Db 20537 EATITGLIQGESEYFRVSAQNEKGISDPRQLSVPIADLVIPPAFKLLFTFTVLAGE 20596
 Qy 502 -----F-----K-----K-- 504
 Db 20597 LKVDVFPFGRPTPAVTHKDNVPLKQTRVNAESTENSSLITIDACREDGVHVVKLTN 20656
 Qy 505 -----A-----MD-----K--G----- 509
 Db 20657 SAGEALETLANVILDKPGPTPGVMDDEVTAISITLWGPXPYDGSINNVIYEKRDTS 20716
 Qy 510 -----AR--A-----I----- 513
 Db 20717 TTTMOIVSATVARTTIKACRLKTGCEYQFRIAENRYKSKSTYLNSEPTVAOYPPKVPBP 20776
 Qy 514 -----M-----VV-----NT--V--N----- 520

Db 20777 GTPVVTLSRDSMEVQWNEBISDGSRGVIGVHLECKERNILMWKLNKTRPIQTKFTTG 20836
 Qy 521 -----Y--Y-----N-----RD-----N-----W----- 527
 Db 20837 LEEGVEYEFVSAENIVIGICKPSKSECVYARDCDPPGRPEALIVTRNSVTLQWKPTY 20896
 Qy 528 -----T-----ELPA--M-----G-----YE----- 536
 Db 20897 DGSKITGYIVEKELEBEGRMKASFTNIIDTHEVGLVEDHRYEFVRVIAENAGVFSE 20956
 Qy 537 -----A--DE-----G-----T-----K--SQ-- 544
 Db 20957 PSESTGAITARDEVDPRIAMDPKYKDTIVVHAGESFKVADIYKPIPTIQWIKGQEL 21016
 Qy 545 -----V-----F--SI-----SG-----D----- 551
 Db 21017 SNTARLEIKSTDFATSLSVDAVARVDSGNVILKAKNAGRSVTVNVKVLDRPGPEGPV 21076
 Qy 552 -----DG-----VK-----L--W----- 557
 Db 21077 VISGVTAECTLAMKPPLODGSDIINVIYERRETSRLVTVVDANQVTLCKTKLEGG 21136
 Qy 558 N-----M--IN-----PD--K-----K----- 565
 Db 21137 NEYFRIMAVNAKYGVEGLESEPVVAKNPVFPDAPAPAPEVTVTKDSMIWVWRPASDG 21196
 Qy 566 --TE--Y-----KR-----N-----N--KE-- 574
 Db 21197 GSEILGYVLEKRDKEGIRWTRCHRLIGELRLVLTGLIEHNDYEFVSAENAGLSESP 21256
 Qy 575 -----D-----F--K--D--KLE--Q--Y----- 583
 Db 21257 PSAYQACDPIYKPGPNRKVIDITRSSVFLSMKPIYDGGCEIQGYIVEKCVSGEW 21316
 Qy 584 -----Y-----PI----- 586
 Db 21317 TMCPTPTGINKNIEVEKLEKEHEYNFRICAINKAGVGHADVGPPIVEKLEAPDIL 21376
 Qy 587 DME-----SF-----N 592
 Db 21377 DLELRKINIRAGSLRLFVPIKGRPTPEVWCKVDEIRDAIIDVTSSTSLVLDNVN 21436
 Qy 593 -----S-----N--K-----P--N-----V--G--D----- 600
 Db 21437 RYDSGKTYLLENSSGTSKSAFYTVRVLDTPSPVNLKVTETIKDSVSIWEPPLDGSK 21496
 Qy 601 -----EK--E-----ID-----F----- 606
 Db 21497 IKNYIVEKREATRKSVAAVVNTNCHKNSWKIDQLOEGSYFVRVLAENEYIGILPAQTADP 21556
 Qy 607 -KFA--P-----D-----T-----D-----K--E----- 615
 Db 21557 IKVAEVPQPPGKITVDVTRNSVLSWTKPEHDGSKIIQYIVEMQAKHSEKMECARVK 21616
 Qy 616 -----L-----Y-----KE--D-----I-----I--V--PA-----G 626
 Db 21617 SLOAVITNLTOGESEYLERVVAVNEKGRSDPRSLAVIIVAKDLYIEPDKPAFSSYVOVG 21676
 Qy 627 -----S-----T--SW--G--P-----RI-----DL--L----- 637
 Db 21677 QDLKIEVPISGRPKPTITWTRKDGFLPKQTRINVTDSLDLTLSIKETHKDDGGQYGTIV 21736
 Qy 638 -----L--KP-----DVSA-----P--G--KN--I--K-- 650
 Db 21737 ANVVGOKTASIEIVTLKPPKPPKGVKFDVSAESITLSNPNPLYTGCGCOITNYIVOKRD 21796
 Qy 651 -----S-----TL-----N-----VI----- 656
 Db 21797 TTTTMDVVSATVARTTLKVTKLKTGTEYQFRIFAENRYGOSFALSDPIVAVOYPIKPEG 21856
 Qy 657 -----N-----G-----K--S--T--Y-----GY 664
 Db 21857 PPGKASNSSECVYARDCDPPGPTEPIWVKRNETTLQWTRKPVYDGSIMTGY 21908

RESULT 8

ABB58144

ID ABB58144 standard; Protein; 7107 AA.

XX AC ABB58144;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 1224.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX OS pharmaceutical.

XX PN Drosophila melanogaster.

XX PD WO200171042-A2.

XX PF 27-SEP-2001.

XX PR 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL02247.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions -

XX PS Disclosure; SEQ ID NO 1224; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
 XX CC useful in developmental biology and in elucidating cell signalling and
 XX CC cell-cell interactions in higher eukaryotes for the development of
 XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
 XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX CC sequences (ABL01840-ABL16175) and the encoded proteins
 XX CC (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 7107 AA;

Query Match 70.4%; Score 3098.2; DB 22; Length 7107;

Best Local Similarity 9.7%; Pred.NO. 6.1e-44;

Matches 579; Conservative 74; Mismatches 10; Indels 5278; Gaps 475;

Qy 1 Y-----P-VV-----L--A-----D-----7

Db 1020 YQFRVIAVNAKGLSPSPDASVPQIVKYKLLKPRIDRSNLKPLIRAGKPIRYDVNVGEP 1079

Qy 8 ----T-----SSSE-----D-A- 14

Db 1080 APVITWYQNDKELKPEELPSSEIKNIPYNTKISIIETVRKHTGIYKIIAVNEHQDEAT 1139

Qy 15 --LNI-----S-----D-----K-E---K-V-A---E-----25

Db 1140 VEVNIIAPPSPRGLDVKDVTKSCKLKWKKPEDDGGKPIYQVERKFDKQGRWVPLG 1199

Qy 26 ----N-----K---EKHE-----NI- 33

Db 1200 RTSANDTEFDVKLQEGHEYQFRVKAINEGESDPLDSDSDSIIAKNPYDAASKPCTPNIV 1259

Qy 34 ----H-----S-A-----M-----E-----TS-----40
 Db 1260 DYNEHWVKLWEAPRSOGAPISGYIIEKKOKFSPIWDEILSTNTSVPEATVEGLVEGNI 1319
 Qy 41 -Q-----D-----F---K-E- 45
 Db 1320 YQFRVRAVNAKGFSDPSDATEPHLAKPRNLKPYINRDWKPKIKVRAGQPVKFDVVKGP 1379
 Qy 46 -----K-----K-----T-----A-----49
 Db 1380 APSLTWFLKETELTSTGOVRLNIDYNTKLTLLDTRKQSGQYKLAENINGVDEAVVEV 1439
 Qy 50 -----K-----VI-----K-----52
 Db 1440 IILDKPSKPEGPIEVSDIHKEGCKLWKPDKDDGIPITGYVIEKMDTATGKWVAGSVD 1499
 Qy 53 -EK--E-----V--V-----S---KNP-----61
 Db 1500 PEKYDIEIKGLDPNHRVQFRVKA VNEEGESEPLETESAITAKNPFDSAPPGLPELEDWD 1559
 Qy 62 --V-----I-D-----NN-----T-S-----68
 Db 1560 EHHVKLWKEPPIRDGSGSPITNYIIEVMDKDSGEFVKA VETDSPVCKGVVKKLEGGQQYKF 1619
 Qy 69 -----N-----EE---AK-----I-K-----E---76
 Db 1620 RVRVNAKAGSDPSEQTWNHVAKPRFLKPHIDRVNLKPVIVKTGLSISLDINIRGEPAPK 1679
 Qy 77 -E---NS-----N-K-----SQ-----83
 Db 1680 VEMFNNSSVTSDSHSVKIDNVYNTKFFVMRAQSSQSKYIIKATNEGEDEAELEVTV 1739
 Qy 84 -----G---DY---TD-----S---89
 Db 1740 LGKPGKPGQLQVNDITKHSCKLWKEKDDGGSPIDYIEIEKLDPHGTQWLPCGKSTEP 1799
 Qy 90 -----P---VNK-----N--TE---NP-----KKE---DK--V 104
 Db 1800 EAKVIGLHGKAYKFRVRAVNAKGESEDELETKPIIAKNPYDEPRPGKPEPTNNDKQDV 1859
 Qy 105 -----V--Y-I-----A-----BF-----110
 Db 1860 DLANDPPKNDGGAPIQKVIQMRDCKSGRAWDSATVPGDKNGTGTGVEEGHEYEFRIVA 1919
 Qy 111 -----K--D-----KE---SGEK---A-IK-----121
 Db 1920 VNAKGPSDPDVSKSVIAKPRFLKPHIDRKNLQKKIMRSGQMLHIDALIKAEPPAKVTWT 1979
 Qy 122 ---EL-SS--LK-N---T---KV---LY---: : : : : 133
 Db 1980 YNKTEIKTSDHIKIEINEDYKTTFIMPVKVRADRGYIIVTAKNDSGSDTVEVELEVLCKPS 2039
 Qy 134 -----T-----Y-----D-- 136
 Db 2040 KPKGPLAVSNVTAETLHLKWEKPEDDGGDPIEQYLVERMDTETGRVPLTTTKTEADVT 2099
 Qy 137 -----R-----I-----F-----N-----140
 Db 2100 GLTEGKEYLFRVKA VNSEGESEPLVTDIPTAKNPFDAADTPGKQPIVDWSGNHCDLKW 2159
 Qy 141 -----G-S-----AIET--TPD-----N-L-----151
 Db 2160 APEDDGGASITGYIVERKDPNTGKQKALETSTPDCKARVNDLIAGNKYQFRIMAVNKAG 2219
 Qy 152 -----D---K--IK--Q---IE--G-----159
 Db 2220 KSKPESDQMTAKDRFAPPKIDRNIKDITIKAGQHIRFDIKVSGEPATKVLHNKAR 2279
 Qy 160 -----IS-----S-----162
 Db 2280 LENDDSNYNIDMESYRTKLTVPISKRPHSGKYTLKAENESGRDEASFVILVDKPGPBG 2339
 Qy 163 -----VE-----R---AQ-----167

Db 2340 PLRVTVHKEGCKLKMNALDGGLPIDHYIIKMDVSGRWLPSCGRFESFAELINDEP 2399
 QY 168 ---K---V-----QPM-----MN-----HA----- 176
 Db 2400 SHEYFRLAVANTEGESEPLETGEQSVIAKNPFDEPGKGTPEAVDMDKDHDVLRPPIN 2459
 QY 177 ---R-----K-EI-----G-----V-----E----- 183
 Db 2460 DGGSPITGVVVEKREKGTDMKIKGEIITPCIGEECKATVPTLNNCEYEFKVAIINAG 2519
 QY 184 -E-----A-ID-----Y-LKS-----IN-----AP----- 195
 Db 2520 PGEPSDASKPITTKPRKLAPKIDRKNI RTYNFKSGEPIFLDINISGEPAPDVIMQNNKS 2579
 QY 196 ---F-----G-K---NF----- 200
 Db 2560 VQTSFSHIENLPYTKYINNPERKDTGLYKISAHNFYQDQVEFOIINTTKPKGPRGP 2639
 QY 201 ---D-----D-----GR-----GMV-- 206
 Db 2640 LEVEVHNDGCKLKKKKPKDGGEPVESYLEKFPDPTGIMLPVGRSDGPEYNDGLPBG 2699
 QY 207 --- 206
 Db 2700 HDYKFRVAVNKEGESEPLETGSIIAKDPFVPTKPGVPEPTDWTANKVELAMPPEASD 2759
 QY 207 ---I----- 207
 Db 2760 GGSPIQYIIVEKDKYSPLEMEKALETNSPTPTATVOGLIEGNEYOFRRVALNKGGLSEPS 2819
 QY 208 -S-----NI-----D-TG----- 213
 Db 2820 DPSKIFTAKPRYLAKIDRKNLRNITLSSGTLKLDANITGEPAPKEWKLSNYHLOGSK 2879
 QY 214 ---T-DY-----R----- 217
 Db 2880 NNTIETPDYTKLVIPTQSRSDGEYLVATNTSGKSVLVNVVITDKPSPPNGPLQISD 2939
 QY 218 -HK----- 219
 Db 2940 VHKEGCHLWKRRPSDDGCTPIEYFOIDKLEPETGCMIPSCRESTEPQVDVTGLSPGNEYKF 2999
 QY 220 ---A-----WR-----ID----- 224
 Db 3000 RVSAVNAEGESQPLVGDESIYAKNPFDEPGKPENIKATDMDKDHVDLAWTPLLIDGSGPI 3059
 QY 225 ---D-----D----- 226
 Db 3060 SCYIIKODKYGKMERALDVPAQCKATIPDLVEGQYKFRVSAVNAAGTGPSPSTPEPI 3119
 QY 227 -AKA-----SM-R-----F-----KKE----- 236
 Db 3120 IAKANNKPPIIDRSSLVEYRIKAGOSFTFDCVSGEPAPQTKMLLKKEKVSXONVKTYN 3179
 QY 237 -D---LK-----G-----T-DK----- 243
 Db 3180 VDYNTKLVNSATRSDSGLYTVFAENANGESADVKYIVIDKRAPNGPLKYDEINSESC 3239
 QY 244 ---NY-----NL-----SD-----KI-----P-H--AF----- 255
 Db 3240 TLHNNPPDDGGQPIIDNYVVEKLDGTGTGRWIPAGETDGVTLAKVGGTLPGHKYFRFRA 3299
 QY 256 -NY-----Y-----N-----GG-KIT-- 264
 Db 3300 KNRQSTSEPLTTAQAIIANPNPDVPTKPTPIKDFKDFVLEWTRPADGSGPITGV 3359
 QY 265 VEK-----Y-----DD-----G----- 271
 Db 3360 VEKDKFSBDMCKEKAISDDITNAHVPLDIEGLKYEFRVAVNAKAGPSPSDATETHVAR 3419
 QY 272 ---R---D-----Y-FD-----P---H-G-M----- 280

Db 3420 PKNTPEKIDRNFMMSDIKIKAGNVFEDVPTGEBLPSCMDTNEGMIINTDRVXISNEDD 3479
 QY 281 ---H-----I-A-----G-----I-L- 286
 Db 3480 RTKIRILDAKRSDTGVYTLTARNINGTDRHNKVTIIDLAPVPBGPLRNGVSKNSIVLR 3539
 QY 287 ---AG-----N-----DTE----- 292
 Db 3540 WRPPKDDGSEITHYVVEKMDNEAMRWVPVGDCTDTEIRADNLIENHDYFRVAVNKG 3599
 QY 293 ---Q-----D-----I----- 295
 Db 3600 QSQPLTTSQPIITAKDPVSHPDKPGOPQATDWKHFVLEWSTPRKDGGAPISSYIIKXP 3659
 QY 296 ---K-----N-----F-----N-G-----I----- 301
 Db 3660 KFGOMERAVALGDNCKAHVPELTNGGEYEFRRVIAVNGSPSDPSSTIICKPRFLAP 3719
 QY 302 ---D-----G-----I-A--P-----NA-----Q----- 309
 Db 3720 FFDKSLNDITVHAGKRLGWTLPLEASPRPLITWLXNGKEIGSNSRGSGLFOHELTFEI 3779
 QY 310 ---I-----F----- 311
 Db 3780 VSLRSDGARYTLILKNEHGSFDSAHAATVLDPRSPPKPLDITKTRDGCHLTWNVPDD 3839
 QY 312 ---S---Y---KM-----YSDAG-S-----G----- 322
 Db 3840 DGGSPILHYIIKNDLSRSTWSADGSMSTHIVHDYTLVRKREYLFRRVAVNAIGESDPLE 3899
 QY 323 ---F---A-G-----D----- 326
 Db 3900 AVNTIIAKNEFDEBPAPKPIITDMDRHDIDLOMAVPSDGAPISEYIIOKKEKSPYW 3959
 QY 327 ---E-T-----M----- 329
 Db 3960 TNVRHVPENKNTTIPILBELTGEQEFYFRIAVANQAGSESPSPDIMAKPRYLPPKITTP 4019
 QY 330 ---FHA---I-E-----D-----SI-----K- 338
 Db 4020 LNEVRIKGLIFHTDHFIEGAPAPATWTLNSNPLSNDRSTISIGHSVHTVNCORS 4079
 QY 339 ---H---N---VD-----VV-----SVS-S-----G--- 349
 Db 4080 DSGIYHLLLRNSSGIDGSEFELVLDLRGPRGEMEBEITANSVTTISMKPKONGSSEI 4139
 QY 350 ---FT-----GT-----G---L--- 355
 Db 4140 SSVYIEKRDLTNKGGMWPAVNVYSAKYNAHVPRLLBECTWYELRMAENLQGRSDPLTSD 4199
 QY 356 ---V---G-E-----K-----Y-----W-- 361
 Db 4200 QPVVAKSQYTVGAPGKPELTDSDKNHITIKMKOPISNGSPPIICYDIERRDVMTGRWIK 4259
 QY 362 ---O-----AI---RALR----- 368
 Db 4260 INGQVPPTAEYQDORVTSNHOYORISAVNAAGNKTSEPSAIFMARPLREKPRYFDGL 4319
 QY 369 ---K-AG-----IP----- 373
 Db 4320 IGRHIVYAGBPVNLNIPISGAPPTTIEWKRGDKLECKRISYETNSERTLFRIDSNR 4379
 QY 374 ---MV-----A----- 377
 Db 4380 RDSGYVTVAANBERGKDTADIEVIVDKPSPPEGPLSTETAPDHISLHWYSPKDDGSD 4439
 QY 378 -TG-----N-----Y-----A---TS----- 384
 Db 4440 ITGYIIETFEFGVDWKPEVGTCPNTFTVKNLVEGKKYVRIRAEINTYGASEALEGRPV 4499
 QY 385 -A-S-----S---S---W---D----- 391
 Db 4500 LAKSPFDPGAPASQPTISAYTPNSANLEMHNPDDCGGKPIINGYIVERREGGEWIKCNVY 4559

Qy	392	- - - L - V A N - - - N - - - H - - - - -	397
Db	4560	: : : P T N T S Y T V S N L R D G A R Y E F R V L A N V A E A G P G H S K Q S D P M T A E H Q R Y R D P P P P K P D R I	4619
Qy	398	- - - L - - - - K - - - M - - - T - - D - - - T - - - G - - -	404
Db	4620	T R N G V T L S W R P R T D G K S R I K G Y V V E M R P K N G K D W K T V N D I P I N S T V Y T V P S L U K E G E E Y S	4679
Qy	405	- - - N - - - - V T - - - - R - - - T - - - A - - -	410
Db	4680	F R V A E N E V G R S D P S Q B P I T I E O P N K C M E L G K V R D I V C R A G D D F S I H V P Y L A P P K P	4739
Qy	411	A - - - - - H - - - - -	412
Db	4740	N A F W S N D N M L D D N R V H K L T T D D A S V V V K N S K R A D S G Q Y R L Q L K N T S G F D T A T I N V R V	4799
Qy	413	- - - E - - - - - D - - A I - - - - A - - - - V A S - - -	420
Db	4800	L D R P S P T R L R A D E F G S D L T L Y M N P P D D G G S A I Q N Y I I E K K E A R S S T W S K V S S F C T V P	4859
Qy	421	- - - A K N - Q - - - T V E - - - - F D K V - N - - - - P K - - -	432
Db	4860	F V R I R N L V N K E Y D F R V I A E N K Y G Q S D P A N T S P B I L A R H E F D - I P N T G I P H G I D S T E D S	4918
Qy	433	- - I - - - - G G - - - E - - - - S - - - - -	439
Db	4919	I T I A W T K P K H D G S P I T G Y I I E K R L L S D D K W T K A V H A L C P D L S C K I P N L I E N A E Y F R V A	4978
Qy	440	- - - Y - - - - - R N - - I - G - A F - - - - -	446
Db	4979	A V N A A G O S A Y S G S S D L I F C R R P P H A P K I T S D L S I R D M T V I A G D E F R I T V P Y H A S P R T A S	5038
Qy	447	- - - F D - - - - K S - K - - - - I T - T - N E - - - - D -	457
Db	5039	W S L N G L E V I P G E R I K F D S N D Y A S W Y N K S A K R D E T G S Y T I T L N N K G S D T A S C H V T V D R	5098
Qy	458	- - G - - - - T - - K A P - - - S - - - - K L - - - K - F V - - Y -	469
Db	5099	P L P Q G P L N A Y D I T P D T C T L A W K T P L D D G S P I T Y V V E K L D N S G S W K V I S S F R W T H Y D	5158
Qy	470	I - G - - - K - - - - G - - - - Q - - - D - - Q - - - D - - - L I	479
Db	5159	V M G L E P H Y K Y N F R V A E N O Y G L S D P L D I E P I V A K H Q F T V P D E P G Q P K V I D M D S G N V T L I	5218
Qy	480	- - - - - G - - - - L - - - - - D - - - - -	482
Db	5219	W T R P L S D G S R I O G Y Q I E Y R D I L D N S W N A Y V I I K D T K Y Q L Y N L I N G S E Y E F R I K A N A	5278
Qy	483	- - - - - L R - - - C K - I - - - A - - - - V - - - - M - D - R I - - -	493
Db	5279	A G L S K P S P S L R F K L K F T V P S P P G A P Q V R V G K N Y V D L K W E K P L R D G G S R I T G Y I I E R	5338
Qy	494	- - - Y T - - - - K - D L K - - - N - - - - -	500
Db	5339	R D I G G A V W K C N D Y N V L D T E Y T V M N L I E M G D Y E F R V P A V N S A G R S P S L C T M P I K V C E V L	5398
Qy	501	- - - A - F - K - - - - K - - - A - - - - M - D - K - - -	508
Db	5399	G G K K P D W I T R L Q D K V A P F G K D Y T L Q C A A S G K P S P T A R W L R N G X E I O M N G R M T C D S K D G V	5458
Qy	509	- - - G - - - A - - - - R - - - - -	511
Db	5459	F R L H I S N V Q G D D G D Y T C E A M N S L G F V N T S G Y L K I G S P P I I N R C P S E L K P L E G D N S K I K I	5518
Qy	512	- - - A I M - - - V - - N - - T - - V N - - - Y - - Y N R - - - D - - -	525
Db	5519	P Y S G Q P L T I L K K N E V I C D S N D D T H V K V N I P D D Y V A I Y I R N I V K S D G G P Y Q I E F T N E S	5578
Qy	526	- - - - - - - - - - N W - - - - -	527
Db	5579	G S A T G E F V H I T G M P S A P T G M W G I S Y I N K N S C M L N W R P P S Y D G G L K V S H Y V I E R K D V S P	5638

Qy	528	-----T-----	-----E-L-P--AM-----	533
Db	5639	HWITVSTCKDTAFNQGLTIENQEIYFRVMVNNENGWGPPLGLEGNIRAKDPIDPPSPG	5698	
Qy	534	-----G-----YE-----A-----D-----	538	
Db	5699	SPOITEIGDFVHLEWKEKPEDSGAHIQGYWIDKREVGSTWQVNVNATICAANQINCL	5758	
Qy	539	EG-----TKS-----Q-----V-----F-----	546	
Db	5759	IEGRQYEFRIFAQNVAGLSTESSASQAVKIIDPOAASPLIVKPLRDANCIONHNAQFTC	5818	
Qy	547	-----S-----IS-----GD-----D--G-----V-----	554	
Db	5819	TINGVPKPTISWYKGAREISNGARYHMYSEGDNHFLINDVGEDADEYVCAVNAKAG	5878	
Qy	555	-----KL-----W-----N-----	558	
Db	5879	STRATLAIMTAPKLVNPPFRDTPAYFDKGVNVKIPFTGFPKPRIHWYRDGENIESGGH	5938	
Qy	559	-----MI-----N-----PDK-----K-----	565	
Db	5939	YTVEVKERHVALIIRGSHLDSGPYRITAENELSGSDTAIIQVOISRDPPRPPPLIESIG	5998	
Qy	566	TE-----VXR-----NN-----KE-DF--	576	
Db	5999	TESLSLWKAPVWDGCSDIITYVERREHPLSSWIRVGNTRFTSMVAVSGLTGPCKEYDFRI	6058	
Qy	577	-----K--D--KL-----	580	
Db	6059	FADNVYGRSDSTSLTIKTESVKKKPIERKWEIDANGRKLKGADGPKDYDSYVFDI	6118	
Qy	581	-----E--Q--Y--Y-----P-----	585	
Db	6119	YSKFPVOPVEISOQSVYDRYDILEEITGAGFVVHRCRERSTGNIPAAKPIPVSHSVEKD	6178	
Qy	586	-----ID-----M--E--S--F-----N	592	
Db	6179	LIRREIDIMNQLHHQKLIINHDAFEDDEMLILEFLSGGELPERITAEGYVMTAEVIN	6238	
Qy	593	-----SN-----KP-----NV-----G-----D--E--K-----	602	
Db	6239	YMRQICEGRHMEQNIHIIIDIKPENIMCOTRSTNVKLIIDFLGATRLDPNEVVKITGT	6298	
Qy	603	-----EI-----DF-----	606	
Db	6299	AEFAAPEVNVNRPVGYTDMWATGVLSYLLSGLSPFAGDNVOTLKNVACDWDQFVDS	6358	
Qy	607	-----K-F-----A-----P-----D-T-----D-----	613	
Db	6359	PKYISEAKOFIRKLLVRNKEKMTAHECLLHPMTGDSAMKQEIINRORYLAYREKLRR	6418	
Qy	614	K-----E-----L--Y--K--E-----D-----I-----	621	
Db	6419	KYEDFERFLLPIGRLSYSLRKLLEKMKYIHDAVDRRQAAPRPFVIRPSSQCFEGQSV	6478	
Qy	622	-----I-V--P-----A-----	625	
Db	6479	KFYCRCIAIATPTLTWSHNNIELRQSVKFMKRYVGDYVFIINRVKLLDRGEYIIAENH	6538	
Qy	626	GS-----T-S-----W-----	630	
Db	6539	YGSREEVVLNVQPLPKEQPRYTESTPVRREPLPYTFWQEBESETAPSTFTLLRPVMO	6598	
Qy	631	-----G--P-----R-----ID-----	635	
Db	6599	ARDCKLLCLSGKPVNVRWYKDGRELSKYEYANTHSDGVVMTWEIIDCKPDSGKYSC	6658	
Qy	636	-----L---LL-----KP-----	640	
Db	6659	ATNCHGTDETCVVIVEGEWTPPEQAHLNFYSGDRKYIEQIPAPLPIVTSROYTS	6718	
Qy	641	-----D--VS-----APG-----	646	

Db 6719 SSVNTSEPGCDKVVNSNSNSGSISSKKVANSLSQAPGSPSRSRATKEILPPDDSLM 6778
 Qy 647 -----K-----I-----K----- 650
 Db 6779 CKPEFTKPLHDLTHDGEQLILTCYKGDPEPQISWNSKSGSLSSDILRLRYKNGIATL 6838
 Qy 651 -----S-----T-----L-----N-----V-----I----- 656
 Db 6839 TINEFPEDEGVITCTATNSVGAETKCKLTIQPLDKINKRKVNAGNARKIVSHLESR 6898
 Qy 657 -----N-----G-----K-----S-----T-----Y-----G----- 663
 Db 6899 FVRDGDVNLACRIIGAQHPDVVWLHNKKEIKPSKDFQYTNEMNIYRLQIAEIFPEDGCT 6958
 Qy 664 Y 664
 Db 6959 Y 6959

RESULT 9

ABG51536 ID ABG51536 standard; Peptide; 5701 AA.

AC ABG51536;
 DT 25-FEB-2003 (first entry)
 XX

Human liver peptide, SEQ ID No 30184.

KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
 KM hypercholesterolaemia; coronary heart disease.
 XX

OS Homo sapiens.

PN MO200157273-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00664.

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488898/53.

PT Human genome-derived single exon nucleic acid probes useful for
 PT analysing gene expression in human adult liver -
 XX

PS Claim 27; SEQ ID No 30184; 658bp; English.

CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
 CC measuring human gene expression in a sample derived from human adult
 CC liver, comprising one of 11109 defined nucleotide sequences given in the
 CC specification (or complements/ fragments). The probe hybridises at high
 CC stringency to a nucleic acid molecule expressed in the human adult
 CC liver. (II) may be used for predicting, measuring and displaying gene
 CC expression in samples derived from human adult liver. The genes
 CC identified may be involved in genetic liver diseases such as cirrhosis,
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
 CC is associated with coronary heart disease. ABG47348-ABG5930 represent
 CC human liver single exon encoded peptides of the invention.
 CC Note: The sequence information for this patent does not appear in the
 CC printed specification but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SO Sequence 5701 AA;

Query Match 69.8%; Score 3070.2; DB 22; Length 5701;
 Best Local Similarity 10.3%; Pred. No. 7e-44;
 Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;

Qy 1 Y-----P---V---LA-D-----T-----S----- 9
 Db 73 YNFRISAINDAGCEPAVIPPVEIVEREMAPDELDELRLRTLVVRAGLSIRIFVPIKGR 132
 Qy 10 -----S-S-----E---D-----A----- 14
 Db 133 PAPEVTWKONINLNKNANIENTESFTLLIIPECNRYDTCKFVWTIENPAGKSGCFVNR 192
 Qy 15 -----LN-----I---S-----D-----K-E----- 21
 Db 193 VLDTPGFLNLRLPDTITKDSVTLHMDPLIDGSRITNYIVEKEATRKSYSTATTKCHK 252
 Qy 22 ---KV-----AEN-----K-E----- 28
 Db 253 CTYKVTGLSBCGEYFFRVMAENEGIGEPTETTEPRVASEAPSPDSLNTMDITKSTVSL 312
 Qy 29 ---KH---E-N---I-----H----- 34
 Db 313 AMPKPKIDGSKITGVYIEAQRKSGDQWTHITTVKGLCEVVRNLTEGEYTFQVMVNSA 372
 Qy 35 ---SA-----M----- 37
 Db 373 GRSAAPRESRPVIVEQMLPELDLRCIYQKLVIAKAGDNIKVEIPVLRPKPTVTKKGD 432
 Qy 38 -----ET-----S-----O----- 41
 Db 433 QILKOTQRVNFEITATSTILINIECVRSDSGPPLTARNIVEGVDTITQVHDIPEGPT 492
 Qy 42 -----DF----- 44
 Db 493 GPIKFEVSSDFVTFPSMDPENOGVPISNVVEMRQTDSTTWELATVIRTYYKATRL 552
 Qy 45 ---E---K-----K-----TAV----- 50
 Db 553 TTGLEQFRVKAQRVGVGPGITSACIVANVPKVPQPTAVTATKOSMTSMHEPL 612
 Qy 51 -----I-----KE-----E---VV----- 57
 Db 613 SDGSPILGVYVERKENGILMOTVSKALVPGNIFKSSGLTDGIAVEFRVIAENMAGSK 672
 Qy 58 -SK-----N-----P-----V---I---D----- 64
 Db 673 PSKPEPMALDPIDPGKFPVPLNITRHTVTLKMAKEFYGGFKITSTIYEKRLPNGRW 732
 Qy 65 -----N---N---T-S-----N-----E-----EA-KIK- 75
 Db 733 LKANFSILNEPFPVSGLTEDAVERFVIAKNAAGAISSPEPBDATCRDDVEAPKIKV 792
 Qy 76 -----E-----E-----N-SNK- 81
 Db 793 DVKEKDTVILKAGAPRLLEADVSGRPPTMEWSKDELBTAKLEIKIADPSTNLVVKD 852
 Qy 82 S-O-GDYT-----D-----S-F----- 90
 Db 853 STRDSCAVYLTATNPGFAKHTINVKVLDPRGPPEGPLAVTEVTSEKCVLSWPPPLDDG 912
 Qy 91 -----V-----N-----K-----N-----T 95
 Db 913 GAKIDHYIVQRETSRLAMTWASEVQVTKLVTKLKNGEYIPRVAAVNVYGVGPLES 972
 Qy 96 E-----N-----PK-----K-----E---DK-----V- 104
 Db 973 EPVLAVPYGPDPDPKXPEVYTTITKDSMVVVCWCHPDGSGSEIINYIVERDKAGORWIK 1032
 Qy 105 -----V-----Y---I-AE-----FK----- 111

Db	1033	CNKKTLTDLRYKVSGLTEGHEYEPRIMAENAGISAPSTSPFYKACDTVFKPGPGNPR	1092
Qy	112	-----D-----K-----	113
Db	1093	VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEDEWQIVTPPAGLKATSYTITGLTE	1152
Qy	114	-----S--GE-----KA-----	119
Db	1153	NOEYKIRIYAMNSEGLGSPALVPGTPKAEDRMLPPEIBLDADLRKVITIRACCTLRFPV	1212
Qy	120	IK-----E-L-----SS-----L-----K-----	127
Db	1213	IKGRPAEKKWARDHGESLDKASTESTSYTLLIVGNVNRFDGKXIIITVENSCKSAF	1272
Qy	128	N-----TK--V-----L-----Y-----T-----Y-----	135
Db	1273	VNVRVLDTPGPPQDLKVKVEVTKSTVLTWDPPLDGGSKIKNYIVKRESTRKAYSTVAT	1332
Qy	136	-----DRI-----F-----N-----G-----S--A-----	143
Db	1333	NCHKTSMKVDQLOEGCSYFRVLAENEYIGLPAETAESVKASERPLPPGKITLMDVTRN	1392
Qy	144	-----I-----I-----E-T-T-----	147
Db	1393	SVLSWEKPEHGGSRILGYIVEMQTKGSDKWATCATVKVTEATITGLIQEESYFRVSA	1452
Qy	148	-----P-----D-----N-----L-----D-K-----I-----	154
Db	1453	QNEKISDPRQLSPVIAKDLVIPAPKLLFNTFTVLAGEDLKVDVDPFGRPTPAVTHK	1512
Qy	155	-----KQ-----	158
Db	1513	DNVPLKQTRVNAESTENSLTTIKDAREDVGHVYVVKLTNSAGAEIETLNVILDKPGP	1572
Qy	159	-----G-----I-----SS-----VE-----R-----A-----	166
Db	1573	PTGPVKMEVTAADSIITLSWGPBKPYDGGSSINNYIVEKRDSTTTTQWIVSATVARTIKAC	1632
Qy	167	-----Q-----K-----V-Q-----P-----M-----M-----N-----	174
Db	1633	RLKTGCEYQPRIAENRYGKSTYLNSEPTVAQYPPKVPKPGPTPVVTLSSRDSMEVQWNE	1692
Qy	175	-----H-ARKE-----I-----GVE-----	183
Db	1693	PISDGSRVIGYHLERKERNISILWYKLNKTIPIOTKFKTTGLEGEVEYFRVSAENIVGI	1752
Qy	184	-----EAI-----	189
Db	1753	GKPSKVSECYVARDPCDPPGRPEALIVTRNSVTLOWKKPTVDGGSKITGYIVEKKELPEG	1812
Qy	190	-----K-S-----	194
Db	1813	RWMKASFTNIIDTHEFTVGLVEDHRYEPRVIAARNAAGVSEPSSESTGAIARDEVDPPRI	1872
Qy	195	-----P-----F-----GK-----N-----	200
Db	1873	SMDPKYKDTIVVHAGESFKVDADYIGKPIPTQWIKGDOELNARLEIKSTDFATSLSV	1932
Qy	201	-----D-----G-----R-----G-MVIS-----	208
Db	1933	KDAVRVDSGNVILKAKNAVAGERSVTNVKVLDRPGPPGPPVIGSVTAETKCTLANWKPLQ	1992
Qy	209	-----N-I-----D-----T-----G-----	213
Db	1993	DGGSIIINYIIVERRETSRLVWTVDANVQTLSCVKYTLLEGNEYTFRIMAVNKYVGEP	2052
Qy	214	-----T-----D-----Y-----	216
Db	2053	ESEPVAKNPFVDPAPKAEVTTVTKDSMIVWNERPADSGGSEITLGVLEKRDKEGIRW	2112
Qy	217	-----R-----H-----KA-----	220
Db	2113	TRCHKRLIGELRLRVTLGIENHDYEFVRVSAENAGLSBSPSPSAQYKACDPYKPGPPNN	2172
Qy	221	-----M-----	221
Db	2173	PKVIDITRSSVFLSWSKPIYDGGCBIOGYIVEKCDVSVGEMWTCTPTGTINKNTINEVEKL	2232
Qy	222	-----RI-----D-D-D--AK-----A--SMR--P	233
Db	2233	LEKHEYNFRICAINKAGVGEHADVPGPPIVEBEKLEAPDIDLDLELRKIINIRAGGSLRUF	2292
Qy	234	-----K-K--E-----D-----L-----K-----GT--	241
Db	2293	VPIKGRPTPEVKWGVKVDGEIRDAIIDVTSSFTSLVDNVRNDYDGGKYTLTLENSGTSK	2352
Qy	242	-----K--D-----KNY-----	245
Db	2353	AFVTVRVLDTPSPVNLKVTEITKDSVSIWEPPLDGGSKIKNYIVEKREATRKSAAV	2412
Qy	246	-----W--L--S-----D-----KI-----	251
Db	2413	VTNCHKNWKIDQLOEGCSYFRVTAENEYIGLPAQTADPIKVAEVPQPGKITVDDVT	2472
Qy	252	-----P-H-----A-----F--	255
Db	2473	RNSVLSWTKPEHDGSGKIIQYIVEMQAKHSEKSECARVKSLOAVITNLTOGEEVLPV	2532
Qy	256	-----N-----Y-----	258
Db	2533	VAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVGDQKIEVPISGRPKPTIW	2592
Qy	259	-----N-----GCK--ITV-----E-----	266
Db	2593	TKDGLPLKQTRINVTDSLDLTLSIKETHKDDGGQYGITVANVVGQKTASIEITLDKP	2652
Qy	267	-----KYDD-----G-----RD-----	273
Db	2653	DPPKGVKFDVDSABESITLSWNPPLYTGCGQITNIVQKRDITTTTVDVVSATVARTILK	2712
Qy	274	-----Y--F--DP-----	277
Db	2713	VTKLKTGYOFRIPAENRYGQSFALSDPIVAQYPPKPGPPGTPPFATAISKDSMVIQW	2772
Qy	278	H-----GMH-----I-----A-----GI-----LA-----	287
Db	2773	HEPVNNGSPVIGYHLERKERNISILWTKNKTIIHDTQFKAQNLBEGIEYEFVRYAENIV	2832
Qy	288	-----G--N-----D--T-----E-Q-----D--I-----K-----	296
Db	2833	GVGKASKNSECYVARDPCDPPCTPEPIMVKNRNEITLQWTKPVYDGGSMITGYIVEKRDLP	2892
Qy	297	-----N-----F-----N-G-I--D--G-I-A-----	305
Db	2893	DGRWMKASFTNVIEQFTVSGLTEDQRYEYFRVIAKNAAGAIKSPSDSTGPIAKDEVELP	2952
Qy	306	-----P-----NA-----	308
Db	2953	RISMDPKFRDITVNVAGETFRLEADVHGKPLPTIEWLRGDKIEBSARCEIKNTDFKALL	3012
Qy	309	-----Q-I-----F-----K-----	311
Db	3013	IVKDAIRIDGGQYILRASNVAGSKFPVNVKVLDRPGPEGPVQVGTVSEKSLTWSPP	3072
Qy	312	-----S-Y-----	316
Db	3073	LQDGGDISHYVVEKRETSRLAWTVVASEVNTLSKVLKLEGEVYFRIMAVNKYVGGE	3132
Qy	317	-----SDAG-----	322
Db	3133	PLESAPVLMKNPFLVLPGPPKSLVETNIAKDSMTVCNWRPDSGGSEIIGYIVEKDRSGI	3192
Qy	323	-----F-----AG-----	330
Db	3193	RWIKCNKRITDLRLRVTLGTEDHEHYEFVRVSAENAGVGEPSPATVYKACDPV-FKPGP	3251

QY 331 ---H-----A-----I-----E-----D-----S-----I 337
 Db 3252 PNNAAIIVDTKNSITLAWCKPIYDGGSELIGVVEICRADEEMQIVTPQICLARTREI 3311
 QY 338 -K--H-----N-----V-----D-----V-----S 345
 Db 3312 SKLTHEOEYKIRVCAIANKVGLGEATSVPGTVKPEBKLEAPBLDSELRKGIIVRAGSA 3371
 QY 346 ---S-----G-----F-----G-----T-----G-----L----- 355
 Db 3372 RIHIFKGRPIPEIITWSREEGEFTDKVOIEKGVNTQISIDNCDRNDAGKYLKLENSSG 3431
 QY 356 ---V-----G-----E-----K-----Y-----W-----A-----I-----R----- 365
 Db 3442 SKSAFVTVKVLDPGPBPNLAVKEVRKDSAPFLWMEPRIIDGAKVKVYIDKRESTRKAY 3491
 QY 366 A-----L-----R-----K-----G-----P-----M-----V----- 375
 Db 3492 ANVSSKSKTSFKVENTLEGALYFRVMAENEFVGVPETDVAKAAEPSPGKVTLLT 3551
 QY 376 -V-----AT-----G-----N-----Y----- 381
 Db 3552 DVSQTSASLWMEKPEHDGSRVLGYVENQPKTEKMSIVAESKVCNAVVTGLSSGOEYO 3611
 QY 382 ---A-----S-----A-----S-----S-----S----- 388
 Db 3612 PRVKAINEKGRDPRVLGVPIVIAKDLTLOPSLKLFPNTYSIOAGEDKLEIFVIGRPRPN 3671
 QY 389 -SW-D--L-----VAN-----N-----HLK-----MT----- 401
 Db 3672 ISWVADGEPLKQOTTRV-INEETATSTVLHIKEGNKDDFGKYTVTATNSAGTLENLSVIV 3730
 QY 402 ---D-----TG-----N-----VT-----R-----TAA-H----- 412
 Db 3721 LEKPGPVGPRFDEVSADPVIISWEPRAYTGCOISNYIVKRDITTTITWMSATVAR 3790
 QY 413 ---E-----D-----A-----IA-----VAS-AK----- 422
 Db 3791 TTIKITLKTGTEYQRIFAENRKYKSAFLDSKAVIYOIFKEPGPPGPTVTSISKDOM 3850
 QY 423 ---N-----Q-----T-----V-----EF-----D-----KV----- 431
 Db 3851 LVQWHEPVNDGGTKIIGYHLEQEKNSILWVKLNTPIODTKFTTGLDEGLEIEFFKSA 3910
 QY 432 -NI-G-GE---S---F-----K-----Y----- 440
 Db 3911 ENIVIGIKPSKVSCEGVARDPCDPGRPEAIVITRNNTLKKWKPAYDGSKITGIYVK 3970
 QY 441 ---R-----NI-----G-----A-----F----- 446
 Db 3971 KDLPGRMKASFTNVLETFVSGLVEDQRYEFRIANNAAGNFSPESSGAILTARDE 4030
 QY 447 ---F-----D-----K-----SK-----I 452
 Db 4031 IDAPNASLDPKXKDVIVVHAGETFVLEADIRCKPIPDVVWMSKDKLELETAAMEIKSTI 4090
 QY 453 -TT-----N-----E-----D-----G-----T-----K----- 460
 Db 4091 QKTTLVVKDCIRTDGOYILKLSNMGTSKSIPIYVAVLDRPGPREGPLKVTVTAEKCYL 4150
 QY 461 A---P-----S-----K-----L-----K-----F-----V----- 468
 Db 4151 ANNPLIODGGANISHYIIEKRETSRLSWTQSTEVQALNYKVTKLLPGNEVIFRMAVAK 4210
 QY 469 Y-IG-----K-----G-----Q-----D-----Q-----D-----LI----- 479
 Db 4211 YGIGBLBSGCVTACNPKPYPGPSPSTPEVSATITKOSMVTWARPVDGSTEIEGYLEKR 4270
 QY 480 ---G-----L-----DLR-----G-----KIA----- 488
 Db 4271 DKEGVMTKCNKKTLLDLRLRVLTGLTEGHSYEFVRVAENAGVGEPSSESVFYRACDALY 4330

QY 489 ---VMD-R-----IY-----TK-----D----- 497
 Db 4331 PPGPSNPKVTDTSSSVSLAWSKPIYDGAIPVKGYVVEVKAADDEWTTCTPTGLGCK 4390
 QY 498 ---LK-NA---F-----KKA-----MD----- 507
 Db 4391 QFTVTKLENTENYFRICAINSEGVEBPATLPGSVVQERIEPPEILDADLRKVVURA 4450
 QY 508 ---KG-----A-----RA-I-----N----- 514
 Db 4451 SATRLFVTIKGRPEPEVKEKAGILTRQOIEVTSFTVLVIDNVTFRPDSGRNLTLE 4510
 QY 515 ---V-----VN----- 517
 Db 4511 NNSGSKTAFVAVRVLDSPSPAVNLTIREVKDVSATLSMEPLIDGAKITNYIVEKRETT 4570
 QY 518 ---TV-----N-----YY-----N-----R 524
 Db 4571 RKAVATTINNCTKTFRIENLOEGCSYFRVLASNEVIGLPAETTEPVKVSBPPLPGR 4630
 QY 525 ---D-----N-----W-----T----- 528
 Db 4631 VTLVDVTRNTATIKKEPESDGSKITGYVVEWOTKSEKMSCTOVKTLLEATTISGLTAG 4690
 QY 529 ---ELP-----A-----MG----- 534
 Db 4691 EBYVRVAANVEKGRSDPRQGVPIVARIIDIKPSVELPHTFNVKAREOLKIDVPFKGR 4750
 QY 535 ---YE-----A----- 537
 Db 4751 POATVNRKDGOTLKETTRVAVSSKTVTSLSIKEASKEDVTELCVSNAGSITVBIT 4810
 QY 538 ---DE-----G----- 540
 Db 4811 IIVDRPGPPGPIRIDEVSCSITTSNMPPEYDGCQISNYIVEKETTSTTHIVSOAV 4870
 QY 541 -T-K-----S-----Q-----V-----FS----- 547
 Db 4871 ARTSIKIVRLTGSBYQFRVCAENRYKSSYSSESAVAEYFPSPGPGTPKVVHATKS 4930
 QY 548 ---I-----SG----- 550
 Db 4931 TMLVTWQPVNDGSGRVIYGHLEYKERSILMSKANKILIIDTQWKGVLDEGLMEYERV 4990
 QY 551 ---D-----DG-----V 554
 Db 4991 YAENIAGIGKSCSKCEPVAPADPCDPGPQPEVNTITRKSVALKWSKPHYDGAKITGYIV 5050
 QY 555 -KL---W-----N----- 558
 Db 5051 ERRELPGDRMLKCNVTNIQETTFEVTLETDORYEFRVPAANMAADVSEPSSESTGPIIVK 5110
 QY 559 ---M-----IN-----P-----D-----KK-----TE----- 567
 Db 5111 DQVEPRVMDVQKFRDVIIVKAGEVLIKINADIAGRPVLIVISWADGIEIEBRARTEIST 5170
 QY 568 ---VK-----R-----N-----N-----K----- 573
 Db 5171 DNHTLLTVKDCIRRDGOYVLTLLKNVAGTRSVAAVCKVLDKPGPAGLEIINGLTAEKCS 5230
 QY 574 ---E-----DFPD-----K-----L-----E-----O-----Y----- 583
 Db 5231 LSWGRPOEDGADI-DYIIVEKREISHLAWTICGELOMTCKYTKLLKGNBYIRRVGV 5289
 QY 584 -Y---P-----I-----D-----M-----ES-----F----- 591
 Db 5280 NKGVGEBLESVAIKALDPFTVPSPPTSLBITSVTKESMTLCWSPESDGSSEIGYIIE 5349
 QY 592 ---NS-----N-----KP-----N-----Y-----G-----D-----E-----K-----E-----I-----D----- 605
 Db 5350 RREKNSILRWVAVNKKPVYDLRVKSTGLREGEYEYRVVAENAGLSLPSSETSPILRABDP 5409
 QY 606 -F-----K-----F-----AP-----DTD-K----- 614

Db 5410 VFLPSPSPKPKIVDSCKTTITIAWPLFDGGAPITGYTVVEYKSDDTDWKTSIQSLRGT 5469
Qy 615 E-----L-----Y-----|||
Db 5470 EYTSISLTTGAEVFRVKNVKGASDPDSDDPQIAKEREERPLFDIDSEMRKTLIVKA 5529
Qy 626 G-S-T-S-----WG-P-----R-----I-----D----- 635
Db 5530 GASFTMTVPFRGRPVFNVLWSKPD TDLTRAYVDTTDSRTSLTIENANRNDGKYTLTIQ 5589
Qy 636 -L-----L-----K-----DV-----SA-----P-----G-----KN-----I----- 649
Db 5590 NVLSAASLTLLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGAPVKNYHIEKREAS 5649
Qy 650 -K-----S-T-----LN-----VIN-----GKSTY-----G-----Y 664
Db 5650 KKAWSVTNNCNRLSYKVTNLQEG-AIYVFRVSGENEF 5686

RESULT 10
ABB36684
ID ABB36684 standard; Peptide; 5701 AA.
XX ABB36684;
AC ABB36684;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #4190 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
OS
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 29319; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5701 AA;

Query Match 69.8%; Score 3070.2; DB 22; Length 5701;
Best Local Similarity 10.3%; Pred. No. 7e-44;

Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;
Qy 1 Y-----P-----VW-----LA-D-----T-----S----- 9
Db 73 YNFRISAINDAGVEPAVDPDVEIVEREMADPFDLAEALRRLTVVRAGLSIRIFVPIKGR 132
Qy 10 -----S-S-----E-----D-----A----- 14
Db 133 PAPEVTWKDINILKNRANIENTESFTLLIPECNRYDTGKFMVTIENPAGKSGFVNVR 192
Qy 15 -----LN-----I-----S-----D-----K-----E----- 21
Db 193 VLDTPGVNLRLPTDITKDSVTLHWDLPIDGGSRIITNIVEKREARKSYSTATTKCHK 252
Qy 22 ---KV-----AEN-----K-----E----- 28
Db 253 CTYKVTGLSEGCYFFRVMANEYIGETETTEPVRKASEAPSPDSLINIMDITKSTVSL 312
Qy 29 -----KH-----E-N-----I-----H----- 34
Db 313 AWPKPKHGGSKITGYVIEAQRKSGDQWTHITTVKGLECVVRNLTGEYEYTFQMAVNNSA 372
Qy 35 --SA-----M----- 37
Db 373 GRSAPRESRPVIVKEQTMLPELDLGIYQKLVIAKAGDNKIVPLGRPKPTVTKKGD 432
Qy 38 -----ET-----S-----O----- 41
Db 433 QILKQTVNFTTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPT 492
Qy 42 -----DF-----K----- 44
Db 493 GPIKFEVSSDFVTFSWDPENDGGVPSINVVEMRQDSTTWVELATTVIRTYKATRL 552
Qy 45 -----E-----K-----K-----TAV----- 50
Db 553 TTGLEVQFRVKAQNRVGVPGITSACIVANVPKVPGPPTQVTAVKDSMTISWHEPL 612
Qy 51 -----I-----KE-----K-----E-----VV----- 57
Db 613 SDGSGPILGYHVERKERNGLWQTVSKALVPGNIFKSSGLTDGIAEYFRVIAENMAGSK 672
Qy 58 -SK-----N-----P-----V-----I-----D----- 64
Db 673 PSKPSPEMLALDPIIDPGKVPPLNITRHTVTLKWAKEPYTGFKITSYIVKRDLPNGRW 732
Qy 65 -----N-----T-S-----N-----E-----EA-KIK- 75
Db 733 LKANFSNILENFTVSGLTEDAAYEPRVIAKNAAGAISSPSEPSDAITCRDDVEAPKIV 792
Qy 76 -----V-----N-----K-----T 95
Db 793 DVKFKOTVILKAGEAPRLEADVSGRPPTMEWSKDGKELEGTALEIKIADFSTNLVNDK 852
Qy 82 S--Q--GDYT-----D-----S-F----- 90
Db 853 STRDSGAYTLTATNPGGFAKHIFNVKLDORPGPEGLAVTEVTSEKCVLSWPPPLDDG 912
Qy 91 -----V-----N-----K-----T 95
Db 913 GAKIDHIVQKRETSRLAWTNVASEVQTKLVTKLKGNEYIFRVMVKNYGVGEPLS 972
Qy 96 E-----N-----PK-----K-----E-DK-----V- 104
Db 973 EPLAVNAPYPPDPKPKNEVTTITKDSMVVVCWGHPSDGGSEIINYIVERDKAGQRWIK 1032
Qy 105 -----V-----Y-----I-AE-----FK----- 111
Db 1033 CNKKTLDLRYKVSGLTEGHEHYEPRIMAENAGISAPSPFPFYKACDTVPKPPGNPR 1092
Qy 112 -----D-----K----- 113
Db 1093 VLDTSRSSISIANKNPIYDGGSEITGYMVEITALEPEDEWQIVTPPAGUKATSYTITGLTE 1152

QY 114 -E-----S-----GE-----KA----- 119
 DB 1153 NQEKIRIYAMNSGLGEPALVPGTPKAEEDMLPPEIELDADLAKVVTIRACCTLRLEFVP 1212
 QY 120 IK-----E-L-----SS-----L-----K--- 127
 DB 1213 IKGRPAPEVKWABDHGSELDCASIESSTSYTLIIVGNVNRFDGSKYLLTVENSSGSKSAF 1272
 QY 128 -N-----TK-V-----L-----Y-----T---Y--- 135
 DB 1273 VNVAVLDTPPPPQDLKKEVTKTSVTLTWDPPLLDGSKIKNYIVEKRESTRKAYSTVAT 1332
 QY 136 -----DRI-----F-----N-----G-----S-A----- 143
 DB 1333 NCHKTSMKVDOLOQEGCSYFFRVLAENEGIGLPAETAESVAKASBRPLPBGKITLMDVTRN 1392
 QY 144 -----I-----I-----E-T-T----- 147
 DB 1393 SVSLSWKEPHEHDGSRILGIYIVEMQTKGSDKMATCATVKTEATITGLIQEGEYSFRVSA 1452
 QY 148 -----P-----D-----N-----L-----D-K-----I----- 154
 DB 1453 QNEKGISDPROLSVPIVIAKDLVIPPAPKLFNTFTVLAGEDLKVDVPFISGPTPAVTWAK 1512
 QY 155 -----KQ-----:-----:-----IE----- 158
 DB 1513 DNVPFLKQTRVNAESTENNSLITIKDACREDVGHVVKLTNSAGEALETTLNVILDKRGP 1572
 QY 159 -G-----I-----SS-----VE-----R---A- 166
 DB 1573 PTGVKMDDEVTAADITLSWGPBPYKDGSSINNYIVEKRDSTTTMQIVSATVARTTIKAC 1632
 QY 167 -----Q-----K-----V-O-----P---M---M---N- 174
 DB 1633 RLKTCGYQFRIAENRYGSKTYLNSEPTVAQYPPKVPGPPTGTVTLSSRDSMEVOMNE 1692
 QY 175 -----H-ARKE-----I-----GVE----- 183
 DB 1693 PISDGSRGVIGYHLERKERNSILMWKLNKTPITPOTKFTTGLEGEVEIEFRVSAENIVGI 1752
 QY 184 -----EA-----D-----Y-----L--- 189
 DB 1753 GKPSKVECYVABDCDPGRPEALIVTRNSVTLQWKKPYTDGSKITGYIVEKKELPEBG 1812
 QY 190 -K-S-----I-----NA----- 194
 DB 1813 RWMKASFTNIIDTHFEVTVGLVEDHRYEERVIARNAAGVFSEPSSESTGAITARDEVDPRI 1872
 QY 195 -P-----F-----GK-----N-----F----- 200
 DB 1873 SMDKYYKDTIVVAHGESFKVDADIYKRPITTIQWIKDQELSTNARLEIKSTDPTSLSV 1932
 QY 201 -----D-----G-----R-----G-MVIS----- 208
 DB 1933 KDAVRVDSGNVILKAKANVAGERSVTYVNVKVLDRPGRPEGPVIVISGVTAEKCTLAMKPELQ 1992
 QY 209 -----N-I-----D-----T-----G--- 213
 DB 1993 DGGSDIINYIVERETSRLVWTVDANVQTLSCVKTLLEGNEYTFRIMAVNKYGVGEPL 2052
 QY 214 -----T-----D-----Y----- 216
 DB 2053 ESEPVAKNPFVVDAPKAPESVTTVTKDSMIVMERPASDGSSELGLVLEKRDKEGIRW 2112
 QY 217 -----R-----H-----KA----- 220
 DB 2113 TRCHKRLIGELRLAVTGLIENHDYEFVRVSAENAGLSEPPSAVQKACDPIYKPGPPNN 2172
 QY 221 -----M----- 221
 DB 2173 PKVIDITRRSSVFLWSKRPIDYGCEIOGYIVEKCDVSGEWTMCTPPTGINKNTIIVEKXL 2232

QY 222 -----RI-----D-D-D-----AK-----A--SNR-F 233
 DB 2233 LEKHEVNFRIICAIINAKAGVEHADVPPIIVEKLEARDIDLDLELRKIIINIRAGSGRLF 2292
 QY 234 -----K-K--E-----D-----L-----K-----GT-- 241
 DB 2293 VPIKGRPTPEVKGKVDGEIRDAIIDVTSSFTSLVDNVRVDSGYTTLTLENSSGTKS 2352
 QY 242 -----KXY-----D-----KXY----- 245
 DB 2253 AFVTVRVLDTPPSPVNLKTEITKDSVSTIWEPELDDGSKIKNYIVEKBEATRKSAAV 2412
 QY 246 -----W-----L-----S-----D-----KI----- 251
 DB 2413 VTNCHKNSWKIDLOQEGCSYFFRVTAENEGIGLPAOTAPRIKAEVQPPRGKITVDVT 2472
 QY 252 -----P-H-----A----- 255
 DB 2473 RNSVLSMTKPEHDGSKIIQYIVEMQAKHSEKSECARVKSLOAVITNLTOGEEYLFVR 2532
 QY 256 -----N-----Y----- 258
 DB 2533 VAVNEKGRSDPRSLAVPIVAKDLVIEBPDVKAPESSYSVQGDUKIEVPISGRKPPTTW 2592
 QY 259 -----N-----GK-ITV-----E----- 266
 DB 2593 TKDGLPLKQTRRINVTSDLSLTTLSIKETHKDDGOGIGIVANVVGOKTASIELVTLDKP 2652
 QY 267 -----KYDD-----G-----RD----- 273
 DB 2653 DPPKPVKFPDVSNAESITLSMNPPLYTGCOITNYIVOKBDTTTWADVAVATFATTLK 2712
 QY 274 -----Y-----F-----DP----- 277
 DB 2713 VTKLKTGYQFRIAENRYGQSFALSDPIVAQYPYKBPCTPPTATIAISKDSWIIQW 2772
 QY 278 H-----GWH-----I-----A-----GI-----LA----- 287
 DB 2773 HEPVNNGSPVIGYHLERKERNSILMTKNAKTTIHDTOFAQNLEBIEIEFRVYAENIV 2852
 QY 288 -G---N-----D---T-----E---O---D---I---K--- 296
 DB 2833 GVGKASKSECYVABDCDPGRPEPIMVKNREITLQWTKPVYDGGSMITGYIVEKDLR 2892
 QY 297 -----N-----F-----N-G-I---D-G-I-A----- 305
 DB 2893 DGRWKKASFTNVLETQFTVSGLTEDORYEFRVIAKNAAGAIKSPSDSTGPTAKDEVELP 2952
 QY 306 -P-----NA----- 308
 DB 2953 RISMDPKFRDITIVVNAGETRLEADVHCKPLPTIEMLRGDKELBESARCEIKXTDFKALL 3032
 QY 309 -----O-I-----F----- 311
 DB 3013 IVKDAIRIDGGQVYILRASNVAGSKSPFVNVKVLDRPGRPEGPVQVGTSEKSLTWSPP 3072
 QY 312 -----S-Y-----K-----M---Y--- 316
 DB 3073 LODGSDISHVVEKRETSRLAMTVVASEVVTNSLKVTKLLEGNEYFRIMAVNKYGVGE 3132
 QY 317 -----SPAG-----SG- 322
 DB 3133 PLBSAPVLMKNPFVLBPGRPKSLEVTNIAKDSMTVYCMNRPDSGSELIIGYIVEKRDNSGI 3192
 QY 323 -----F-----AG-----DETMF----- 330
 DB 3193 RMIKCNKRITDLRLAVTGLTEDHEHYEFVRVSAENAGVGEPSPATVYVYKACDPV-FKRG 3251
 QY 331 -----H-----A-----I---E---D-----S---I 337
 DB 3352 PTNAHIVDTTKNSITLWAKGRIYDGGSELIGYVEICKABEEMQIVTPTGTLVATRFEI 3311
 QY 338 -K--H-----N-----V-----D-----VV---S- 345

Db 3312 SKLTEHQEYKIRVCALNKVGLGEATSVPGTVRPEDKLEAPELDLDSSELKGIWVRAGGSA 3371
QY 346 -----VS-S---G-FT-----G---T-----G---L----- 355
Db 3372 RIHIFPKGRPTTEITWSREEGEFTDKVOIEKGWVYQTLSIDNCDRNDACKYILKLENSSG 3431
QY 356 -----V---G-----E-K---Y-WQ-----A-----I-----R----- 365
Db 3432 SKSAFVTVKVLDPGPPQNLAVKVRKDSAFVWEPPIIDGGAKNVYIDKRESTRKAY 3491
QY 366 A-----L-----R-KA-----GIP-----MV----- 375
Db 3492 ANVSSKCSKTSKVENLTGEGAIYFRVMAENBEFGVGPVETVDVAKAABPPPPGKVTLT 3551
QY 376 -V---AT-----G-----N-----Y- 381
Db 3552 DVSQTSASLWMEKPEHDGGSVGLGVVEMQPKGTEKWSIVAESKVCNAVVTGLSSQGEYQ 3611
QY 382 -----A-T-----S-----S-----S----- 388
Db 3612 FRKAYNEKGSDDRVLGVPVIAKDLTIQPSLKLPTNTYSIOAGEDLKIEIPVIGRPREN 3671
QY 389 -SM--D--L---VAN-----N--HLK-----MT----- 401
Db 3672 ISWVKDGEPLKQTRV- NVEETATSTVLHIKEGNKDDPGKYTVTATNSAGTATENLSVIV 3730
QY 402 -----D-----TG-----N-VT--R---TAA-H----- 412
Db 3731 LEKPPVPVGRFDEVSADFVISMPEPAYTGCOISNVIVEKRDTTTTTHMVSAVAR 3790
QY 413 -----E-----D--A-IA-----VAS-AK----- 422
Db 3791 TTITKITKTGETYOFRIFAENRYGKSAPLDSKAVIVQYFPKEPPGPPFTVTSIKDQM 3850
QY 423 -----N-----Q-----T-V---EF-----KV-- 431
Db 3851 LVQWHEPVNDGKTKIIGYHLEQEKNSILWVKNLKTPIQDTKFKTTGLDEGLEYEKVS 3910
QY 432 -NI-G-GE---S--F-----K-----Y----- 440
Db 3911 ENIVGKPSKVSCEFVARPCDPPGRPEAIVITRNNVTLKWKPKPAYDGSKITGIVIEK 3970
QY 441 -----R-----NI-----G-----A--F----- 446
Db 3971 KDLPGRMWKASFTNVLETFEFTVSLGVEDQRYEFRVIAARNAAGNFSEPSDSSGAITARDE 4030
QY 447 -----F---D--K-----SK-----I 452
Db 4031 IDAPNASLDPKYKDVIVVHAGETFVLEADIRGKPIPDVWWSKDGKELEBETAARMEIKSTI 4090
QY 453 --TT-----N-----E--D-----G-T--K--- 460
Db 4091 QXTLLVVKDCIRTDGQYILKLSNVGGTKSIPITVKVLDRLPGPPEGLKVTGTAEKCVL 4150
QY 461 A--P-----S---K-----L---K-----F---V--- 468
Db 4151 ANWPPLQDGGANISHYIIEKRETSRLSWTQVSTEQALNKYKVTLLPGNEYIFRVMAVVK 4210
QY 469 Y-IG-----K--G---Q-----D-----Q--D-----LI----- 479
Db 4211 YGIGLESPGSPVACNPKYKPPPPSTPEVSAITKDSMVVTWARPVDDGGTEIEGYILEKR 4270
QY 480 -----G-----L--DLR-----G---KIA----- 488
Db 4271 DKEGVWTKCNKKTLLDLRLVLTGLTEGHSYEFVRVAENAAAGVGEPSVVFYRACDALY 4330
QY 489 -----VMD--R-----IY-----TK-----D----- 497
Db 4331 PRGPPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKYGVVVEVKEAAADEWTCTPTTGLQK 4390
QY 498 -----LK-NA---F-----KKA-----MD----- 507

Db 4391 QFTVTKLKENTENFRICAINSEGVEGPATLPGSVVAQBERIPPEIELDADLRKVVVLRA 4450
QY 508 -----KG-----A-----RA-I-----M----- 514
Db 4451 SATLRLFTVIKGRPEPEVKWEKAEGILTDRAQIEVTSSTFMTLVIDNVTRFDSGRYNLTLE 4510
QY 515 -----V-----VN----- 517
Db 4511 NNSGKTAFVNVRLDPSAPVNLITREVKDSVTLSEPPIDGAKITNIVEKRETT 4570
QY 518 -----TV-----N-----YY-----N-----R 524
Db 4571 RKAYATIINNCTKTTFRLENLQEGCSYFRVLASNEYGIGLPAETTEPVKVPSEPLPPGR 4630
QY 525 ---D--N-----W-----T----- 528
Db 4631 VTLVDVTRNTATIKWEKPESDGGSKITGVVEMQTKGSEKWTCTQVKTLTATISGLTAG 4690
QY 529 -----ELP-----A-----MG- 534
Db 4691 BEYVFRVAAVNEKGRSDRQLGVPVIAARDIEIKPSVELPFHTFNVKARBQLKIDVPFKGR 4750
QY 535 -----YE-----A----- 537
Db 4751 POATVNRKDGOTLAKETTRVNVSSKTVTSLSIKEASKEDVGTVELCVSNSAGSITVPIT 4810
QY 538 -----DE-----G----- 540
Db 4811 IIVLDRPGPGPIRIDEVSCDSTISWNPPEYDGGCQISNYIVEKETTSTTWHIVSOAV 4870
QY 541 --T--K-----S--Q-----V---FS----- 547
Db 4871 ARTSIKIVRLTGSYQFRVCAENRYGKSSYSESAVVAEYFPSPGPGTKPVVHATKS 4930
QY 548 -----I-----SG----- 550
Db 4931 TMLVTWQVPVNDGGSRVIGYHLEYKERSILWKSANKILLIADTQMKVSLDGLMEYEVYV 4990
QY 551 -----D-----DG-----V 554
Db 4991 YAENIAGIKCSKCEPVPARPCDPPGQPEVTNITRKSLSLWSPKPHYDGGAKITGYIV 5050
QY 555 ---KL---W-----N----- 558
Db 5051 ERRELDPGRWLKCNVTNIQETVFEVTELTEDORYEFRVARNAAADSVSPSSSTGPIIVK 5110
QY 559 -----M-----IN-----P-----D---KK--TE--- 567
Db 5111 DDVEPRVMDVKFRDVIWVKAQVVKINADIAGRPLPVISWAKOGIEIEERARTIIST 5170
QY 568 -----VK-----R--N-----SG-----N---K--- 573
Db 5171 DNHTLLTVKDCIRRDGTQYVLTKNVAGTRSVAVACKVLDKPPGAPGLEINGLTAEKCS 5230
QY 574 -----E---DFKD---K---L-----E-Q-----Y----- 583
Db 5231 LSWGRPOEDGGADI-DYIYVEKRETSHLAWTICEGELQMTCKVTKLLKNEYIFRVTGV 5289
QY 584 --Y---P---I--D-----M-----ES-----F--- 591
Db 5290 NKYGVGEPLSVAIKALDPFTVPSPPTSLEITSVTKESMTLWSPESPESGGSEISGYIIE 5349
QY 592 ---NS-----N-KP--N--V-----G-D-E---K-----E---I--D- 605
Db 5350 RREKNSLRWVRNKKPVYDLRVKSTGLREGCEYEVYRVAENAAAGLSLSPSETPLIARADP 5409
QY 606 -F-----K--F---AP-----DTD-K----- 614
Db 5410 VFLPSPPKPKIVDSGKTTITIAWVKPULPDGGAPITGYTVYKKSDDTDWKTSIOSLRGT 5469
QY 615 E---L---Y-----KE-----DI-----IVPA 625
Db 5470 EYTISGLTTGAEYVFRVSKVNVKASDPSDSDPQIAKEREEREPFLDIDSEMRKTLIVKA 5529

QY 114 -E-----S-GE-----KA----- 119
Db 1153 NOEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPEIEDLADLRKVTVIRACCTLRFLVP 1212
QY 120 IK-----E-L-----SS-----L-----K----- 127
Db 1213 IKGRPAPEVKWARDHGESLDKASIESTSYLLIIVGNVNRFDGKYILTVENSSEKSAF 1272
QY 128 -N-----TK--V-----L-----Y-----T-----Y----- 135
Db 1273 VNVRLDTPGPPQDLKVAEVTKTSVTLTWDPPLDGGKIKNYIIVEKRESTRKAYSTVAT 1332
QY 136 -----DRI-----F-----N-----G-----S-----A----- 143
Db 1333 NCHKTSWKVDLQEGCSYFVRVLAENEYIGLPAETASVKASERPLPGKITLMDVTRN 1392
QY 144 -----I-----E-----T-----T----- 147
Db 1393 SVLSWEKPEHGGSRILGYIVEMQTKGSDKWATCATVKVTEATITGLIQGEEYSFRVSA 1452
QY 148 -----P-----D-----N-----L-----D-----K-----I----- 154
Db 1453 QNEKISDPROLSVPVIAKDLVIPPAFKLLFTFTVLAGEDLKVDPVPTIGRPTPAVTWHK 1512
QY 155 -----KQ-----IE----- 158
Db 1513 DNVPLKQTRVNAESTENNSLLTIKDACREDVGHVYVVKLTNSAGEAIELTLNVILDKPGP 1572
QY 159 -G-----I-----SS-----VE-----R-----A----- 166
Db 1573 PTGPVKMBEVTAADSTLSWGPCKYDGGSSINNYIIVEKRDSTTTTQWIVSATVARTIKAC 1632
QY 167 -----Q-----K-----V-----Q-----P-----M-----M-----N----- 174
Db 1633 RLKTGCEYQFRIAENRYGKSTYLNSEPTVAQYPEKVPGPPTVVTLSSRDSMEVQWNE 1692
QY 175 -----H-ARKE-----I-----GVE----- 183
Db 1693 PISDGGSRVIGYHLERKERNLSLWVKNKTPIPQTKFTGLGEBGEYEFVRVSAENI VGI 1752
QY 184 -----EAI-----D-----Y-----L----- 189
Db 1753 GKPSKVSCEYVARDPCDPGRPEALIVTRNSVTLQWKKPTYDGGSKITCYIIVEKKELPEG 1812
QY 190 -----K-S-----I-----NA----- 194
Db 1813 RWMKASFTNIIDTHPEVTGLVEDHRYEFVRVIAARNAAGVPSEPSSESTGALTARDEVDPRI 1872
QY 195 -----P-----F-----GK-----N-----F----- 200
Db 1873 SMDPKYKOTIVVHAGESFKVDADIYGKPIPTQIWIKGQELSENTARLEIKSTDFATSUSV 1932
QY 201 -----D-----G-----R-----G-MVIS----- 208
Db 1933 KDAVRVDSGNYILKAKNVAGERSVTNVVKVLDRCPPGPEPVVISGVTAEKCTLANKPPLO 1992
QY 209 -----N-I-----D-----T-----G----- 213
Db 1993 DGGSDIINIVERRRETSLRVMTVDANVQTLSCVKTKLEGNETFRIMAVNKYGVGEPL 2052
QY 214 -----T-----D-----Y----- 216
Db 2053 ESEPVVAKNPFVVPAPKAPETVTTVKOSMIVVWERPASDGGSEILGVYLEKRDKEGIRW 2112
QY 217 -----R-----H-----KA----- 220
Db 2113 TRCHKELICELRLVTGLIENHDYBFRVSAENAGLSEPSPSAYQKACDPYKPGPPNN 2172
QY 221 -----M----- 221
Db 2173 PKVIDITRSSVFLSWSKPIYDGGCEIQGYIVEKDVSVGEWMTCTPPTGINKNTINEVEKL 2232
QY 222 -----RI-----D-D-D-----AK-----A-SMR-F 233

Db 2233 LEKHEYNFRICAINKAGVGEHADVPGPPIIVEKLEAPDIDLLELRKIINIRAGCSLRFP 2292
QY 234 -----K--K--E-----D-----L-----K-----GT----- 241
Db 2293 VPIKGRPTPEVKWGVGEIRDAALIDVTSSFTSLVDNVNRYDSGKYTLTLENSSGTKS 2352
QY 242 -----N-----D-----KNY----- 245
Db 2353 AFVTVRLDTPSPPPVNLKVTBITKDSVITWEPLDGGSKIKNYIVEKREATRKSAAV 2412
QY 246 -----W-----L-----S-----D-----KI----- 251
Db 2413 VTNCHKNSWKIDQLQEGCSYFVRVTAENEYIGLPAQTADPIKVAEVPQPPCKIIVDDVT 2472
QY 252 -----P-H-----A-----F----- 255
Db 2473 RNSVLSWTKPEHDDGSKIIQIVIVEMQAKHSEKARVKSLOQAVITNLTOGEBYLFV 2532
QY 256 -----N-----Y----- 258
Db 2533 VAVNEKGRSDPRSLAVPIVAKDLVIEPDKPAPSSYSVQVGDULKIEVPISGRPKPTTW 2592
QY 259 -----N-----GGK--ITV-----E----- 266
Db 2593 TKDGLPLKQTRINVTDSLDLTLSIKETHKDDGGQYGITVANVVGOKTASIEIIVLDKP 2652
QY 267 -----KYDD-----G-----RD----- 273
Db 2653 DPPKGPVKFDDVSAESITLSWNPPLYTGCGQITNYIVQKRDTTITVWDVWSATVARTLTK 2712
QY 274 -----Y--P--DP----- 277
Db 2713 VTKLKTGYOFRIFAENRYGOSFALES DPIVAQYPYKEPGRPFPFATAISKDSMTIQW 2772
QY 278 H-----GMH-----I-----A-----GI-----LA----- 287
Db 2773 HEPYNNGGSPVIGYHLERKERNLSILWTKVNTIHTDTQFKAQNLBEGIEYEFVRVAENIV 2832
QY 288 -G-----N-----D--T-----E--Q-----D-----I-----K----- 296
Db 2833 GVGKASKNSECVARDPCDPPTPEPIMVKRNEITLQWTKPVYDGGSMITGIVIVEKRDLP 2892
QY 297 -----N-----F-----N--G-I--D--G-I-A----- 305
Db 2893 DGRMKASFTNVITQFTVSGLTEDQRYEFVRVIAKNAAGAIKSPSDSTGPITAKDEVLP 2952
QY 306 -----P-----NA----- 308
Db 2953 RISMDPKFRDTIVVNAGETFRLEADVHGKPLPTIENLRGDKIEESARCEIKNTDFKALL 3012
QY 309 -----Q-I-----F----- 311
Db 3013 IVKDAIRIDGGQYILRASNVAGSKFPVNVKVLDRPPEGPVQVGTVTSEKSLTWSPP 3072
QY 312 -----S-Y-----K-----M-----Y----- 316
Db 3073 LQDGGSDISHYVVEKRETSRLAWTVVASEVVTNSLKVTKLEGNBYVFRIMAVNKYGVGE 3132
QY 317 -----SDAG-----SG----- 322
Db 3133 PLESAPVLMKNPFFVLPGPPKSLLEVNTNIAKDSMTVCWNRPSDGGSEIIGYIVEKDRSGI 3192
QY 323 -----F-----AG-----DETWF----- 330
Db 3193 RWIKCKRRTIDRLRVLTGLTEDHEYEYFRVSAENAAAGVGEPSPATVYVYKACDPV-FKPGP 3251
QY 331 -----H-----A-----I-----E-----D-----S-----I 337
Db 3252 PTNAHIVDTTKNSITLAWGKPIYDGGSEILGVYVVEICKADEEEOIIVTPQTGLRVTRFBI 3311
QY 338 -K--H-----N-----V-----D-----VW-----S- 345

Db 3312 SKLTHEOBYKIRVCAIANKVGLGEATSVBGTVPKPEDKLEAPBLDLSBLRKGIVVRAGSSA 3371
 Qy 346 -S-S--G-FT-----G--T-----G-----L----- 355
 Db 3372 RIHIFPKGRPELTWSHEBEFTDKVOIEKGVNTQLSINCDRNDGAKYILKLENSG 3431
 Qy 356 -V-----G-----E--K-----Y-WQ-----A-----I-----R----- 365
 Db 3432 SKSAFVTVKVLDTGPPONLAVKEVRKOSAFVWEPPIIDGAKVKNVVIDKRESTRKAY 3491
 Qy 366 A-----L-----R-KA-----GIP-----MV----- 375
 Db 3492 ANVSSKSKTSFKVENLTEGAIYYFRVAENEFVGVEVETVDAVKAAPSPSPGKYLT 3551
 Qy 376 -V-----AT-----G-----N-----Y- 381
 Db 3552 DVSQTSASLWMEKEPBHGGSRVLGVVEMQPKTEKMSIVAESKYCNAVVTGLSSGGEYQ 3611
 Qy 382 -A-T-----S-----A-----S-----S----- 388
 Db 3612 FRVKAAYNKGKSDPRVLGVPVIAKDLTIQPSLKLPFNYSIOAGBDLKEIPVIGRRPN 3671
 Qy 389 -SW--D--L-----VAN-----N--HLK-----MT----- 401
 Db 3672 ISWVDGEPKQOTTRV--NVEETATSTVLHIKEGNKDDFGKYTVATNAGTATENLSIV 3730
 Qy 402 -D-----TG-----N-VT--R--TAA--H----- 412
 Db 3731 LEKGPVGVPRPEVSADPVAVISWEPAYTGGCOISNYIVEKDDTTTTHMVASATVAR 3790
 Qy 413 -E-----D--A-IA-----VAS-AK--- 422
 Db 3791 TTIKITKUKTGEVQFRIFAENRYKSAAPLDSKANIVQYPRKEPBGPTPVTSISKQOM 3850
 Qy 423 -N-----Q-----T-V--EF-----D-----KV-- 431
 Db 3851 LVQWHEPVDGDTKILGYHLEQEKNSILWKNLNTPIQTRKFTTGLDEGLEVEFKYSA 3910
 Qy 432 -NI--G-GB--S--F-----K--Y--- 440
 Db 3911 ENIVIGIKRPSKVSCEGFVARDCDPPRPEAIVITRNNTLKKKRAYDGSKITGYIEK 3970
 Qy 441 -R-----NI-----G-----A--F----- 446
 Db 3971 KDLDPGRMKKASFTNVLETFEFTVSGLVEDQRYERFVIRMAAGNSEPSDSSGATARDE 4030
 Qy 447 -F-----D--K-----SK-----I 452
 Db 4031 IDAPNASLDPKYKQVIVVHAGETFLVLEADIRGKPIPDVWMSKGELEBETARMEIKSTI 4090
 Qy 453 -TT-----N-----E--D-----G-T--K-- 460
 Db 4091 QKTTLVVKDCIRTDGQYILKLSNVGTSKSIPIVTKVLDPRGPPGPKVTGVTAEKCYL 4150
 Qy 461 A--P-----S-----K-----L-----K-----F--V-- 468
 Db 4151 AMNPLODGCANISIIYEKETSRLSMTQVSTEQALNKKYTKLLPGENEYIFRMAVANK 4210
 Qy 469 Y-IG-----K--G--Q-----D-----Q--D-----LI--- 479
 Db 4211 YGIGPLESGPVACNPKPPGPPSTPEVSATIKDSMVTVARPVDDGTEIEGYILEKR 4270
 Qy 480 -G-----L--DLR-----G-----KLA----- 488
 Db 4271 DKEGRWTKCNKKTLLTLRLRVLTGLTEGHSYEFVLAENAAGVGPSPSVFYRACALY 4330
 Qy 489 -VMD--R-----Y-----TK-----D----- 497
 Db 4331 PRGPPSNKVTDTSSSVSLAMSKPIYDGGAPVKGVVVEKAADEMTTCTPTGLOGK 4390
 Qy 498 -LK--NA--F-----KKA-----MD----- 507
 Db 4391 QFTVTKLKENTEYNFRICAINSEGVGEBATILPGSVVAQERIEPPEIILDADLRKVVLRA 4450

Qy 508 -KG-----A-----Ra-I-----M----- 514
 Db 4451 SATILRFVTIKGRPEPVKMEKAGILTDRAQIEVTSFTMLVINDVTRPDSGRNLTLE 4510
 Qy 515 -V-----VN----- 517
 Db 4511 NNSGKTAFVNVVRVLDSPAPVNLTIREVKKDSVTLSEWPEPLIDGAKITNYIVEKRETT 4570
 Qy 518 -TV-----N-----YY-----N----- 524
 Db 4571 RKAVATTNNCTKTTPRIENLQEGCSYFVRVLSNEXYIGLPAETPEPVKSEBPLPGR 4630
 Qy 525 -D--N-----W-----T----- 528
 Db 4631 VTLVDVTRNATTIKMEKPEBDSGSKITGVVENMOTKSEKSKTCTQVKTLEATISGLTAG 4690
 Qy 529 -ELP-----A-----MG- 534
 Db 4691 BEYFRVAAVNEKGRSDPQOLGVPIARDIEIKPSVELPFHTFNVKAREQLKIDVPFKGR 4750
 Qy 535 -YE-----A----- 537
 Db 4751 POATVNRKDGQTLKETTRVNVSSSKTVTSLSIKASKEDVGTYELCVSNAGSITVPIT 4810
 Qy 538 -DE-----G----- 540
 Db 4811 IIVUDRPGPPGPIRIDEVSCDSITISMNPPYDGGCOISNYIVEKETTSTTHHIVSOAV 4870
 Qy 541 -T--K-----S--Q-----V--FS----- 547
 Db 4871 ARTSIKIVRLTTSGEYQFRVCAENRYGKSSYSESSAVAAVAFSPGPPGPKVVAHTKS 4930
 Qy 548 -I-----SG----- 550
 Db 4931 TMLVQVPVNDGSRVIGYHLEEKERSSILMSKANKILADTQMKVSGDGLMEYERV 4990
 Qy 551 -D-----DG-----V 554
 Db 4991 YAENIAGIKCSKSCBPVPAPDCDPPGQPEVTNITRSVSLKMSKPHYDGAKITGYIV 5050
 Qy 555 -KL-----W-----N----- 558
 Db 5051 ERRLLPGRWLKKNYTNIGETFEVTELTEDQRYERFVIRMANADSVSEPSGPIVX 5110
 Qy 559 -N-----IN-----P-----D--K--TE----- 567
 Db 5111 DVEPPRVMDVKFERDVIIVVAGEVLKINADIAGRPLPVISWADGIEIEBRARTEIIST 5170
 Qy 568 -VK-----R--N-----N-----K-- 573
 Db 5171 DNHTLLTVKDCIRRDQGYVLTLLKNVAGTNSVAANKVLDKPPGAGLEITNGLTAEXCS 5230
 Qy 574 -E--DKD--K-----L-----E--Q-----Y----- 583
 Db 5231 LSWGRPEDGADI--DYIYEKRETSHLAWTICGELQMTSCVTKLLKNGEYIFRVYGV 5289
 Qy 584 -Y--P-----I--D-----M-----BS-----F-- 591
 Db 5290 NKYGVGEPLSEVAIKALDPFTVPSPPTSLBITSYTKESMTLWMSRPSDGSSEISGYIIE 5349
 Qy 592 -NS-----N--KP--N--V-----G--D--E-----K-----E--I--D-- 605
 Db 5350 RREKNSLRWVRVNNKPYVLDLKVKSTGRREGCEYERYVAENAGLSLPSSTPLIRADP 5409
 Qy 606 -F-----K--F--AP-----DTD--K----- 614
 Db 5410 VFLPSPSKPIVDSKTTITIAVVKPLFDGCAIITGYTVBYKKSDDTDKMTSISQSLRG 5469
 Qy 615 E--L-----Y-----KE-----DI-----IYPA 625
 Db 5470 EYTSIGLTGAELYVFRVSVNKGASDPDSDDPOIAKEREERBPLFDIDSEMRKTLIVKA 5529

QY 626 G-S-T-S-----WG-P-----R-I-----D----- 635
Db 5530 GASFTWTFGRGPRVPVNLWSKPDLDLRTRAVDTTDSRTSLTIENARNDSGKVTLTIQ 5589
QY 636 -L-L--K-----P-----DV--SA-----G--KN-I----- 649
Db 5590 NVLSAASLTLLVVKVLDTPGPTNITVQDVTKESAVLSWDVPENDGGAPVKNVHIKREAS 5649
QY 650 -K--S-T--LN-----VIN-----GKSTY-----G--Y 664
Db 5650 KKAWSVTNNCNRLSYKVTNLQEG-AIYFRVSGENEF 5686

RESULT 12
AAM57448
ID AAM57448 standard; Protein; 5701 AA.
XX AC AAM57448;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29553.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human brains -
Example 4; SEQ ID NO: 29553; 650pp + Sequence Listing; English.
PS The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX Sequence 5701 AA;

Query Match 69.8%; Score 3070.2; DB 22; Length 5701;
Best Local Similarity 10.3%; Pred. No. 7e-44;
Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;
QY 1 Y-----P-VV--LA-D-----T-----S----- 9
Db 73 YNFRISAINDAGVGPAVIPDEIVEVERMAPDFELDAELRRLTVVRAGLSIRIFVPIKGR 132
QY 10 -----S-S-----E---D-----A----- 14

Db 133 PAPEVTTKDNILNKNRANIENTESFTLLIIPECNRYDTGKFMVTIENPAGKKSQGVNVR 192
QY 15 -----LN-----I--S-----D-----K-E----- 21
Db 193 VLDTPGPVNLNRPDITKDSVTLHWDLPIDGSGRITNYIVEKREATRSYSTATTCKHK 252
QY 22 --KV-----AEN-----K-E----- 28
Db 253 CTYKVTGLSEGEYFFRMAENEYIGEPTEETTEPVKASEAPSPDSLINIMDITKSTVSL 312
QY 29 -----KH-----E-N--I-----H----- 34
Db 313 AWPKPKHGGSKITGYVIEAQRKSGDQWTHITTVKGLCVVRNLTEGEYTFQWMAVNSA 372
QY 35 --SA-----M----- 37
Db 373 GRSAPRESRPVIVKEQTMPELDLRGIYQKLVIKAGDNIKVEIPVLRPKPTVTWKGD 432
QY 38 -----ET-----S-----Q----- 41
Db 433 QILKQTVNFETTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPT 492
QY 42 -----DF-----K----- 44
Db 493 GPIKFEVSSDFVTFSDPPENDGGVPISVYVEMRQDSTTWVELATTVIRTTTKATRL 552
QY 45 --E--K-----K-----TAV----- 50
Db 553 TTGLEYPQRVKAQNRVGVGPGITSACIVANYPFKVPGPGTPOVTAVTKDSMTISWHEPL 612
QY 51 -----I-----KE-----K-----E-VV----- 57
Db 613 SDGSPILGYHVERKERNGILMQTVSKALVPGNIFKSSGLTDGIAEYFRVIAENMAGSK 672
QY 58 -SK-----N-----P-----V--I--D----- 64
Db 673 PSKSEPMALDPIPPGKPVPLNITRHTVTLKWKAPETTGFKITSIVEKRDLPNGRW 732
QY 65 -----N--N--T-S-----N-----E-----EA-KIK- 75
Db 733 LKANFSNILENEFTVSGLTEDAAVEFRVIAKNAAGAIAPPSEPSDAITCRDDVEAPKIV 792
QY 76 -----N-----E-----N-SNK- 81
Db 793 DVKFKDVTILKAGEAFRLADVSGRPPTMBSKDGKEGTAKLEIKIADPSTNLVNDK 852
QY 82 S--Q--GDYT-----D-----S-F----- 90
Db 853 STRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPEPLAVTEVTSEKCVLSWFFPLDDG 912
QY 91 -----V-----N-----K-----N-----T 95
Db 913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLTKLLKGNEYIFRMAVKNKYGVGPLES 972
QY 96 E-----N-----PK-----K-----E-DK-----V- 104
Db 973 EPVLAVNYPGPPDPKPNPEVTTITKDSMVVCGHPDSDGGSEIINYIVERDKAGQRIK 1032
QY 105 -----V-----Y--I-AE-----PK----- 111
Db 1033 CNKKTLDRLRYKVSGLTEGHEVEFRIMAENAGISAPSTSPFFYKACDVFKPGPGNPR 1092
QY 112 -----D-----K----- 113
Db 1093 VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEDEWQIVTPPAGLKATSYTITGLTE 1152
QY 114 --E-----S--GE-----KA----- 119
Db 1153 NOEKIRIYAMNSEGLGEPALVPGTPKAEEDRLMPPPEILDADLRKVTIRACCTLRLFVP 1212
QY 120 IK-----E-L-----SS-----L-----K----- 127

Db 1213 IKGRPAPEVKWARDHGESLDKASIESTSYTLIIIVGNVRPDGKYILTVENSSGSKSAF 1272
 Qy 128 -N-----TK-V-----L-----Y-----T---Y----- 135
 Db 1273 VNVRLDTPGPPDLKKEVTKTSVTLTWDPPLDGGSKIKNIYVEKRESTRKAIYSTVAT 1332
 Qy 136 -----DRI-----F-----N-----G-----S---A----- 143
 Db 1333 NCHTSMWKVQDLOEGCSYFRVLAENEYIGLPAPETASVKASERPLPGKITLMDVTRN 1392
 Qy 144 -----I-----E-T-T----- 147
 Db 1393 SVSLSWEKPEHDGSRILGIYIVEMQTGSKDMKATCATVKEATITGLIOGEEYSFRVSA 1452
 Qy 148 -----P-----D-----N-----L-----D-K-----I----- 154
 Db 1453 ONEKISDPROLSPVIAKDLVIPPAFKLFTFTVLAGEDKLDKDVPIGRPTPAVTWK 1512
 Qy 155 -----KQ-----I-----E----- 158
 Db 1513 DNVPLKQTRVNAESTENNSLLTIKDACREDVGHVVKLTNSAGAIETLNVIVLDKRP 1572
 Qy 159 -G-----I-----SS-----VE-----R---A- 166
 Db 1573 PTGPKMDEVTAADITLSWGPCKYDGGSSINNIYIEKRDSTTTWQIVSATVARTTIKAC 1632
 Qy 167 -----Q-----K-----Y-Q-----P---M-----M---N- 174
 Db 1633 RLKTGCEYQFRIAEENRYGKSTYLNSEPTVAQYEPKVGPGPTPTLSSRDSMEVQNE 1692
 Qy 175 -----H--ARKE-----I-----GVE----- 183
 Db 1693 PISDGSNVIGVHLERKERNILMWKLNKTPIPOTKFTTGLEBEVEIEFRVSAENIYGI 1752
 Qy 184 -----EAI-----D-----Y-----L--- 189
 Db 1753 GKPEKSECYVARDPCDPGPREALIVTRNSVTLQMKKPYDGGSKITGYIVEKKELBEG 1812
 Qy 190 -K-S-----I--NA----- 194
 Db 1813 RWMKASFTNIIDTHEVTGLVEDHRYEPRVIARNAAGVFSPESEGTGAITARDEVDPRI 1872
 Qy 195 -P-----F-----GK-----N-----F----- 200
 Db 1873 SMDEKYKDTIVNHAGESFKVDADYKRPITTIQWIKGQELSNTRALEIKSTDPAISLSV 1932
 Qy 201 -D-----G-----R-----G-MVIS----- 208
 Db 1933 KDAVRVDGNYILAKKNVAGERSVTNVNVKVLDRPGPREGPVVISGVTAEKCTLAMKPEPLQ 1992
 Qy 209 -N-I-----D-----T-----G--- 213
 Db 1993 DGGSDIINYIVERETSRLVWTVVDANVQTLSCVKTLLEGNEYFRIMAVNKYGVGEPL 2052
 Qy 214 -----T-----D-----Y----- 216
 Db 2053 ESEPVNAKPPVNDAPKAPAEVTTYTKDSMTIWMERPASDGSSELGLVEKROKEGIRW 2112
 Qy 217 -----R-----H-----KA----- 220
 Db 2113 TRCHKRLIGELRLKRTGLIEHNDYEFRVSAENAGLSEPPSAVQKACDPIYKGPNN 2172
 Qy 221 -----M----- 221
 Db 2173 PKVIDITRSSVFLWSKPIYDGCCEIOGYIVEKCDVSGEWTMCTPPTGINKTNIYEKL 2232
 Qy 222 -----RI-----D-D-D-AK-----A-SM-F 233
 Db 2233 LEKHEYNRIKAIKAKGAGEHADVGPPIIVEBEKLEAPDIDDLERKIIINRAGSGLP 2292
 Qy 234 -----K-K-E-----D-----L-----K-----GT-- 241
 Db 2293 VPIKGRPTPEVWKGVDEIRDAIIVTSSFTSLVLNDVNRVDSGKYTLTLENSSGTKS 2352

Qy 242 -----KNY----- 245
 Db 2253 AFVTVRVLDTPSPPVNLKYTEITKDSVITWEPPLDGGSKIKNIYVEKREATRKSAAV 2412
 Qy 246 -----W--L--S-----D-----KI----- 251
 Db 2413 VTCHKNISWKIDQOEGCSYFRVTAENEYIGLPQATADPIKVAEVPQPGKITLVDDVT 2472
 Qy 252 -----P-H-----A-----F-- 255
 Db 2473 RNSVLSWTKEHDGSKIIQYIVEMQAKHSEKNSCARVKSLOAVITNLTOGEEYLFV 2532
 Qy 256 -N-----Y----- 258
 Db 2533 VAANEKGRSDPRSLAVPIVAKDLIEPDVKAPESSYSVQGDLEIYPISGRPKITTW 2592
 Qy 259 -----N-----GK-ITV-----E----- 266
 Db 2593 TKDGLPLKQTRIVNTDSDLTTLTIKETHKDDGGGYGIVANVVGOKTASIEIVTLDKP 2652
 Qy 267 -----KYDD-----G-----RD----- 273
 Db 2653 DPKGPKVFDVSAESITLSWNPPLYTGGCOITNYIVQKRDTTTTVADVVSATVARTTLK 2712
 Qy 274 -----Y--F--DP----- 277
 Db 2713 VTKLKTGYQFRIFAENRYGQSFALSDPIVAQYPKGPGPTPRATAISKDSMWIQW 2772
 Qy 278 H-----GKH-----I-----A-----GI-----LA--- 287
 Db 2773 HEPVNGGSPYIGVHLERKERNISLMTKANKTIIHDPOFAQNLEBGEIEFRVYAENIV 2832
 Qy 288 -G--N-----D--T-----E--Q--D--I-----K--- 296
 Db 2833 GVGKASKSECVARDCDPGTPBEPIMVKRNEITLQWTKPVYDGGSMITGYIVEKRDLP 2892
 Qy 297 -----N-----F-----N--G-I--D--G-I-A----- 305
 Db 2893 DGRMMKASFTNVIETQFTVSGLTEDOREYEFPRVIAKANAAGAIKSPSDSTGPITAKDEVELP 2952
 Qy 306 -P-----NA----- 308
 Db 2953 RISMDPKFROTIYVNAGETPRLFADVHGKPLPTIEMLRGDEIESARCEIKNTDFKALL 3012
 Qy 309 -----Q-I-----F----- 311
 Db 3013 IVKDAIRIDGQYILRASNVAGSKSPVNVKVLDRPGPREGPVQVTGTSKSLTWSPP 3072
 Qy 312 -----S-Y-----K-----M---Y--- 316
 Db 3073 LODGSDISHYVEKRETSRLAMTVVASEVNTSLKVTKLEGNERYFRIMAVNKYGVGE 3132
 Qy 317 -----SDAG-----SG- 322
 Db 3133 PLESAPVLMKNPFLVLPGRPKSLEVTNIAKDSMTVCMNRPDSDGSSEIIGYIVEKRDNSGI 3192
 Qy 323 -----F-----AC-----DETMF----- 330
 Db 3193 RMIKCNKRITDLRLKRTGLTEHDHEYFRVSAENAGVGEPSATVYVYKACDPV-FKGP 3251
 Qy 331 -H-----A-----I-----E--D-----S---I 337
 Db 3352 PTNAHIVDTTKNSITLMAKGRPIYDGSSEILGYYVEICKADEEWOIYTPQTLRLVTRPEI 3311
 Qy 338 -K--H-----N-----Y-----D-----VV-----S- 345
 Db 3312 SKLTHQBYKIRVCAINKVGLGENTSVPGTIVKPEDKLEAPBLDLSLRKQIVVRAGGSA 3371
 Qy 346 -----VS-S--G-FT-----G--T-----G-----L----- 355
 Db 3372 RIHIFKGRPTPELTWREGEFTDKVQIEKGAVVYTOISIDNCRNDAGKYILKLENSSG 3431

QY 356 -V--G--E-K--Y-WO--A--J--R-- 365
Db 3432 SKSAFTVVKVLDTPGPQNLAVKEVRKDSAFVWEPPIIDGGAKVKVYVVDKRESTRKAY 3491
QY 366 A--L--R-KA--GIP--MV-- 375
Db 3492 ANVSSKCKTSFKVENLTCGAIYYFRVMAENFGVGVETVDVAKAAEPSPGKVTLT 3551
QY 376 -V--AT--G--N--Y- 381
Db 3552 DVSOVSASLMWEKPEHDDGSRVLGVVEMOPKGTCKMSIVAESKVCNAVVTGLSSGOEQY 3611
QY 382 -A-T--S--A--S--S-S-- 388
Db 3612 FRVAYNEKSKDPRVLGVPIAKDLTIQPSLKLPHNTYSIQAGEDLKIPIVIGRPRPN 3671
QY 389 -SW--D--L--VAN--N--HLK--MT-- 401
Db 3672 ISHWKGEPLKQTRV-NVEETATSTVHIKEGKNKDDFKYTVTATNSAGTATENLSVIV 3730
QY 402 -D--D--TG--N-VT--R--TAA-H-- 412
Db 3731 LEKPGPVGVPRDEVSADFWISWEPAYTGGQISNYIVEKRDTTTTTHMVVSATVAR 3790
QY 413 -E--D--A-IA--V--AK-- 422
Db 3791 TTIKITLKTGTBYQFRIFAENRYGKSAPLDSKAVIOYPPKPPGTPPVTISKDQM 3850
QY 423 -N--N--Q--T-V--EF--D--KV-- 431
Db 3851 LVQWHEPVNDGKTIQYHLEQEKESILWLKLNKTIQDTKFKTGLDGLGEYFVKVA 3910
QY 432 -NI-G-E--S-F--K--Y-- 440
Db 3911 ENIVGIGKPSKVCSECFVARDCPPGPRPEAIVITRNNVTLKWKPAYDGGSKITGYIVEK 3970
QY 441 -R--NI--G--A-F-- 446
Db 3971 KDLPGRMKASFTNLETFVSGLVEDORYEPRVIRNAGNFSEPSDSSGAITARDE 4030
QY 447 -F--D--K--SK--I 452
Db 4031 IDAPNASLDPKYKDVVHAGETVLEADIRGKPIPDVWMSKDGKLEETAARMEIKSTI 4090
QY 453 -TT--N--E--D--G-T--K-- 460
Db 4091 QKTTLVVKDCIRTDGGQYILKLSNVGKTSIPITVKVLDLRGPPGPKVTGVTAEKCYL 4150
QY 461 A--P--S--K--L--K--F--V-- 468
Db 4151 AMNPLODGGANISHYIIEKRETSRLSWTQVSTEQALNYKVTKLLPGNEYIFRMAVVK 4210
QY 469 Y-IG--K-G--Q--D--Q--D--L-I-- 479
Db 4211 YGIGEPLSGVTCACNYKPPGPPSTPEVSAITKDSMVVTWARVDDGGTEIGYLEKR 4270
QY 480 -G--L-DLR--G--KIA-- 488
Db 4271 DKEGVRMTCNKKTLTDLRLRVTLGTEGHSYEFRAAENAGVGPSEPSVYFACDALY 4330
QY 489 -VMD--R--Y--TK--D-- 497
Db 4331 PPGPPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKGVYVEVKEAAADEWTTCTPPTGLQK 4390
QY 498 -LK-NA--F--KKA--MD-- 507
Db 4391 QFTVTKLNTENYFRICAINSEGVPATLPGSVVAQERIEPPEIELDLARKVVLRA 4450
QY 508 -KG--A--RA-I--M-- 514
Db 4451 SATLRLFTVIKGRPEVKEKABGILTDRQIEVTSFTMLVIDNVTRFDSGRYNTLE 4510
QY 515 -V--VN-- 517

Db 4511 NNSGSKTAFVNVRLDSPSAPVNLTIREVKKDSVTLSWEPPLIDGGAKITWIVEKRETT 4570
QY 518 -TV--N--Y--N--R 524
Db 4571 RKAVATITNNCTKTTFRIENLOEGCSYFVRVLASNEYIGLPAETTEPVKVSEPLPPOR 4630
QY 525 -D--N--W--T-- 528
Db 4631 VTLVDVTRNTATIKWKPESDGGSKITGVVEMOTKGEKWKSTCTQVKTLEATISGLTAG 4690
QY 529 -E--ELP--A--MG- 534
Db 4691 EBYVFRVAANVEKGRSDRQLGVPIARDIBIKPSVELPPHTFNVKAREQLKIDVPFKGR 4750
QY 535 -YE--A-- 537
Db 4751 POATVNRKDGQTLKETTRVNVSSKTVTSLISKEASKEDVGTVELCVSNAGSITVPIT 4810
QY 538 -DE--G-- 540
Db 4811 IIVLDRPGPGPIRIDEVSCDITISWPPPYDGGCQISNVIVEKETTSTTHWIVSOAV 4870
QY 541 -T--K--S--Q--V--FS-- 547
Db 4871 ARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSPGPGTKPVVHATKS 4930
QY 548 -I--SG-- 550
Db 4931 TMLVTWQVPVNDGGRVIGVHLEYKERSILLWSKANKILIAQTQMKVSGLDGLMYEYRV 4990
QY 551 -D--DG-- 554
Db 4991 YAENIAGIGKSKCEPVPARDCPPGQPEVTNITRKSLSKWSKPHYDGGAKITGIV 5050
QY 555 -KL--W--N-- 558
Db 5051 ERRELPGRMKCNVNIQETYFEVTELTEDQRYEFRVFNAAADSSESTGPIIVK 5110
QY 559 -M--IN--P--D--KK--TE-- 567
Db 5111 DDVEPPRVMMDKFRDVIIVKAGEVLKINADIAGRPLPVIISWAKDGBIEERARTEIIST 5170
QY 568 -VK--R--N--N--K-- 573
Db 5171 DNHTLLTVKDCIRDTGQYVLTKNVAGTRSVAVNCKVLDKPPGPPAGPLEINGLTAEKCS 5230
QY 574 -E--DFKD--K--L--E-Q--Y-- 583
Db 5231 LSWGRPQEDGGADI-DYIVIVEKRETSHLAWTICEGELQMTCKVTKLLKNEYIFRVTGV 5289
QY 584 -Y--P--I--D--M--ES--F-- 591
Db 5290 NKYGVEPLESVAIKALDPFTVPSPPTSLEITSVTKESMTLCSRPESDGGSEISGYIIE 5349
QY 592 -NS--N-KP--N--V--G-D-E--K--E--I--D- 605
Db 5350 RREKNSLRWVRNKKPVYDLRVKSTGLREGCEYRVRVAENAGLSLSETSPLTRAEDP 5409
QY 606 -F--K--F--AP--DTD-K-- 614
Db 5410 VFLSPSPKPIVDGSKTITITIAWVKPLFDGAPITGYTVVEYKSDDDTDWKTISIQLRGT 5469
QY 615 E--L--Y--KE--DI--IVPA 625
Db 5470 EYTTISGLTTGAEYVFRVKS VNKVGASDPDSDPOIAKEREBEPLFDIDSEMRKTLIVKA 5529
QY 626 G-S-T-S--WG-P--R--I--D-- 635
Db 5530 GASFTMTVPFRGRPVNVLWSKPDTLRTRAVVDTTDSRTSLTIENANRNDGSKYTLTQ 5589
QY 636 -L--L--K--P--DV--SA--G--KN--I-- 649

Db 5590 NWLSASITLVVKKVLDTPGPPNTITVQDVTKESAVLSNDVENDGAEVKNYHIEKREAS 5649
 QY 650 -K---S-T--LN-----VIN---GKSTY-----G---Y 664
 Db 5550 KKAWSVTNNCNRSLSKYKTNLOEG-AIYFRVSGENER 5686

RESULT 13
 AAM69843
 ID AAM69843 standard; Protein; 5701 AA.

XX AAM69843;
 AC
 XX
 DT 06-NOV-2001 (first entry)
 XX

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30149.

XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukemia; lymphoma; myeloma.

XX Homo sapiens.

XX W0200157276-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

PT Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 30149; 658bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukemia and myeloma. The present sequence is a

CC protein encoded by one of the probes of the invention.

XX

SO Sequence 5701 AA;

Query March 69.8%; Score 3070.2; DB 22; Length 5701;

Best Local Similarity 10.3%; Fred. NO. 7e-44; Indels 4958; Gaps 495;

Matches 578; Conservative 68; Mismatches 14;

QY 1 Y-----P--VV--LA-D-----T-----S-----9

Db 73 YNFRISAINDAGVGEPAVIFDVEIVERMADPFELDALRKLTVVRAGLSIRIFVPIKGR 132

QY 10 -----S-S-----E--D-----A-----14

Db 133 PAPERVTKONINLKNRANIENTESFTLLIPECNRVDTGKFWMTIENPAGKSGFVNV 192

QY 15 -----LN-----I--S-----D-----K-E-----21

Db 193 VLDTPGVLNLRPTDITKDSVTLHMDLPLIDGSRITNYIVEKREARTRKSYSTATTKCHK 252

QY 22 ---KV-----AEN-----K--E-----28

Db 253 CTYKVTGLSEGCCEFFFRMAENEGIGEPTETTEPVKASEAPSPDLSINIMDTKSTVSL 312

QY 29 -----KH-----E-N--I-----H-----34

Db 313 AMPKPKHDSGSKITGVVIEAQRKSGDQWTHITTVKGLCECVVRNLTEGEYTFQVMAVNSA 372

QY 35 --SA-----M-----37

Db 373 GRSAPRESRPVYKEQTMLELDLRLGYOKLVIAKADNIKVEIPVLGRPKPTVMKKGD 432

QY 38 -----ET-----S-----Q-----41

Db 433 QILKQTORVNETTATSTILININECVRSDSGPVPLTARNIVGEVDVITIQVHDI PGPPPT 492

QY 42 -----DF-----K-----44

Db 493 GPIKFEVSSDFVTFSPMDPPENDGVPIISVYVEMROTDSTVWELATTVIRITYKATRL 552

QY 45 -----E-----K-----TAV-----50

Db 553 TTGLEIYQFRYKAQRVGVGPGITSACTIVANYPFKVPGPGTPOVTAVTXDSMTISWHEPL 612

QY 51 -----I-----KE-----K-----E--VV-----57

Db 613 SDGSPILGYHVERKENGLIMQTVSKALVPGNIFKSSGLTDGIAYEFPRVIAENMACKSK 672

QY 58 -SK-----N-----P-----V-----I-----D-----64

Db 673 PSKPSBPMALDPIDPPGKVPPLNTHVTYTLKWKAKBEYTGFKITSYIVEKRLPNGRW 732

QY 65 -----N--N--T-S-----N-----E-----EA-KIX-75

Db 733 LKANFSNILENEFTVSLTEDDAVEFRVIAKNAAGAISSPSEPSDAITCRDDVEAPKIKV 792

QY 76 -----E-----E-----E-----N-SNK-81

Db 793 DVKFKDVTILKAGEAFRLBADVSGRPPTMWSKDGLEGTAKLEIKIADFSTNLVNDK 852

QY 82 S--Q--GDYT-----D-----S-F-----90

Db 853 STRDSGAYTLTATNPGFAKHIFNVKVLDRPGPPEGLAVTEVTSEKCVLSWPPPLDDG 912

QY 91 -----V-----N-----K-----N-----T95

Db 913 GAKIDHYIVQKRETSRLAMTNVASEVOVTKLVTKLKMEYIFRMAVNVKVGCPLES 972

QY 96 E-----N-----PK-----K-----E-DK-----V-104

Db 973 EPVLAVNYPGPRPPPKRPEVTTITKDSMVVCMGHPDSDGSEIINYIERDKAGORWIK 1032

QY 105 -----V-----Y--I-AB-----PK-----111

Db 1033 CNKKTLTDLRYKVSGLTEGHEYEFRIMAENAGISAPSPTSPFYKACDVTYKPPGPNR 1092

QY 112 -----D-----K-----113

Db 1093 VLDTSRSSISIAMNPKPIYDGSSEITGYMEIALPEEDEMOI VTPPAGUKATSYITGLTE 1152

QY 114 -E-----S--GE-----KA-----119

Db 1153 NQEKIRIYAMNSGCGEPALVPOTPKAEDRMLPEPEILDADLRKVTIRACTLRFLVP 1212

QY 120 IK-----E-L-----SS-----L-----K--127

Db 1213 IKGRPAEVMWARDHGSGLDKASIESTSYTLTLVGNVNRFPDSGYILITVENSSGSKSAF 1272

QY 128 -N-----TK--V-----L-----Y-----T--Y-----135

Db 1273 VNVKVLDTGPPGPDLLKXEVYKTSVTLTWDPPLLDGSKIKNYIVEKRESTRKAYSTVAT 1332

QY 136 -----DRI-----F-----N--G-----S--A-----143

Db 1333 NCHKTSKMDQLQEGCSYFVRVLAENEYIGLPAETAESVKASBRPLPPGKITLMDVTRN 1392
Qy 144 -----I-----E-T-T----- 147
Db 1393 SVSLSEKPEHGGSRILGYIVEMQTKGSKWATCATVKVTEATITGLIOGEEYSFRVSA 1452
Qy 148 -----P-D-D-N-L-D-K-----I----- 154
Db 1453 QNEKIGSDPROLSPVPIAKDLVIPAPKLLFNTFTVLAGEDLVDPVPIGRPTPAVTWHK 1512
Qy 155 -----KQ-----IE----- 158
Db 1513 DNVPLKQTRVNAESTENNSLLTIKDACREDVGHVVKLTNSAGEAETLANVLVDKPGP 1572
Qy 159 -G-----I-----SS-----VE-----R-----A- 166
Db 1573 PTGPVKMDEVADGITLSWGPYPKDGSSINNYIVEKRDSTTTWQIVSATVARTTIKAC 1632
Qy 167 -----Q-----K-----V-Q-----P-M-----M-----N- 174
Db 1633 RLKTGCEYQFRIAENRYGKSTYLNSEPTVAQYFPKVPKPGPGTVPVTLSSRDSMEVQWNE 1692
Qy 175 -----H-ARKE-----I-----GVE----- 183
Db 1693 PISDGSRVIGYHLERKERNLSILWKLNKTPIQTKFKTTGLEGEVEYEFVRVSAENIVGI 1752
Qy 184 -----EAI-----D-----Y-----L----- 189
Db 1753 GKPSKVSCEYVARDPCPPGRPEALIVTRNSVTLOWKKPTYDGSKITGYIVEKKELPEG 1812
Qy 190 -K-S-----I-----NA----- 194
Db 1813 RWMKASFTNIIDTHFEVTGLVEDHRYEPRVIRNAAGVFPSEPSSESTGAIARDEVDPRI 1872
Qy 195 -P-----F-----GK-----N-----F----- 200
Db 1873 SMDPKYKDTIVVHAGESFKVDADIYGPPIQTIQWIKGQELSNARLEIKSTDFAATLSV 1932
Qy 201 -----D-----G-----R-----G-MVIS----- 208
Db 1933 KDAVRVDSGNYILKAKNVAGERSVTNVVKVLDKRGPPGPPVIGVTAETKCTLAWKPLQ 1992
Qy 209 -----N-I-----D-----T-----G----- 213
Db 1993 DGSDDIINYIVERRETSRLVTVVDANVOTLSCKVTKLLENEYTFRIMAVNKYGVGEPL 2052
Qy 214 -----T-----D-----Y----- 216
Db 2053 ESEPVAKNPFVDPAPKAPETVITKDSMIVVWERPASDGSSEILGYVLEKRDKEGIRW 2112
Qy 217 -----R-----H-----KA----- 220
Db 2113 TRCHKRLIGELRLVGTGLIENHDEYFRVSAENAGLSEPPSPAYQKACDPIYKPGPN 2172
Qy 221 -----M----- 221
Db 2173 PKVIDITRSSVFLSWSKPIYDGGCEIOGYIVEKCDVSGEWTCTPPTGINKNTNIEVKL 2232
Qy 222 -----RI-----D-D-D-AK-----A-SMR-F 233
Db 2233 LEKHEYNFRICAINKAGVGEHADVPPIIVEKLEAPDIDLDLELRKIINIRAGGSLRF 2292
Qy 234 -----K-K-E-----D-----L-----K-----GT-- 241
Db 2293 VPIKGRPTPEVKWGKVDGEIRDAALIDVTSSFTSLVDNVNRYDSGKYTLTLENSGTSK 2352
Qy 242 -----D-----KNV----- 245
Db 2353 AFVTVRVLDTPSPPNVNLKVTETKDSVITWEPPLLDGGSKIKNYIVEKREATRKSAAV 2412
Qy 246 -----W-----L-----S-----D-----KI----- 251
Db 2413 VTNCHKNSWKIDQLQEGCSYFVRVTAENEYIGLPAQADPIKVAEVPQPPGKITVDVT 2472

Qy 252 -----P-H-----A-----F----- 255
Db 2473 RNSVLSLWTKPEHGGSKIIQYIVEMQAKHSEKARVKSLOAVITNLTGEEYLFVR 2532
Qy 256 -----N-----Y----- 258
Db 2533 VAVNEKGRSDPRSLAVPIKADLVIEPDVKAPFSSYSVQVGQDLKIEVPISGRPKPTIW 2592
Qy 259 -----N-----GK-----ITV-----E----- 266
Db 2593 TKDGLPLKQTRINVTDSLDTLTSIKETHKDDGGQYGITVANVVGQKTASIEIVTLDKP 2652
Qy 267 -----KYDD-----G-----RD----- 273
Db 2653 DPPKGPVKFDDVSAESITLSWNPPLYTGCGQITNIVOKRDTTTTVDVVSATVARTTLK 2712
Qy 274 -----Y-----F-----DP----- 277
Db 2713 VTKLKTGTEYQFRIFAENRYGQSFALSDPIVAQYYPKEPGPGTFFATAISKDSMVIQW 2772
Qy 278 H-----GMH-----I-----A-----GI-----LA----- 287
Db 2773 HEPVNNGGSPVIGYHLERKERNLSILWTKVNTIIHDTQFKAQNLBEGIEYEFVRVAENIV 2832
Qy 288 -G-N-----D-T-----E-Q-----D-I-----K----- 296
Db 2833 GVGKASKNSECYVARDPCDPPGTPPEPIMVKRNEITLQWTKPVYDGGSMITGVIVEKRDLP 2892
Qy 297 -----N-----F-----N-G-I-----D-G-I-A----- 305
Db 2893 DGRWMAKSTNVVETQFTVSGLTEDQRYEPRVIAKNAAGAIKSPKSDSTGPIAKDEVELP 2952
Qy 306 -----P-----NA----- 308
Db 2953 RISMDPKFRDITVNVNAGETFRLEADVHGKPLPTIEWLRGDKIEESARCEIKNTDFKALL 3012
Qy 309 -----Q-I-----F----- 311
Db 3013 IVKDAIRIDGGQYILRASNVAGSKSPVNVKVLDRPGPEGPVQVGTVTSEKCSLTWSP 3072
Qy 312 -----S-Y-----K-----M-----Y----- 316
Db 3073 LODGSDISHYVVEKRETSRLAWTVVASEVNTSLKVLKLEGGNEVPRIMAVNKYGVGE 3132
Qy 317 -----SDAG-----SG- 322
Db 3133 PLESAPVLMKNPFLPGPPKSLVETNIAKDSMTVCNRPDSGSGSEIIGYIVEKDRSGI 3192
Qy 323 -----F-----AG-----DETWF----- 330
Db 3193 RWIKCNKRRTDLRLRVLTGLTEDHEYEFRVSAENAGVGEPSPATVYVYKACDPV-FKPGP 3251
Qy 331 -----H-----A-----I-----E-----D-----S-----I 337
Db 3252 PTNAHIVDTTKNSITLANGKPIYDGSSEILGVVEICKADEBEWQIVTPQTLGRVTRPFI 3311
Qy 338 -K-H-----N-----V-----D-----V-----S- 345
Db 3312 SKLTEHQYKIRVCALNKVGLGEATSPGTVKPEDKLEAPDLDELSELKGVIVRAGGSA 3371
Qy 346 -----VS-S-G-FT-----G-T-----L----- 355
Db 3372 RIHPFKGRPTPEITWRSREGEFTDKVOIEKGVNVTQLSIDNCDNRNDAGKYILKLENSSG 3431
Qy 356 -----V-----C-----E-K-----Y-WQ-----A-----I-----R----- 365
Db 3432 SKSAFVTVKVLDTFGPPQNLAVKEVRKDSAFVWEPFIIDGAKVKNYVIDKRESTRKAY 3491
Qy 366 A-----L-----R-KA-----GIP-----MV----- 375
Db 3492 ANVSSKCSKTSFKVENLTGEGAIYFVRVMAENEFVGVPVETVDVAKAAEPSPPGKVTLT 3551


```

QY 376 -V--AT-----G-----N-----Y- 381
Db 3552 DVSQTSALMWEKPEHDSGRVLGVVEMQKTEKMSIVAESKVCNANVTGLSSGGQYQ 3611
QY 382 -A-T--S-----A-----S-S----- 388
Db 3612 FRVAAYNEKGSDBRVLGVPVIAKDLTIQPSLKLPFNTYSIOAGBDLKEIPVIGRPRPN 3671
QY 369 -SW--D--L-----VAN-----N-HLK-----MT----- 401
Db 3672 ISWAKDEPLKQOTTRV-NVEETAISTVLHIKEGNKDFGKYTATATNSAGTATENLSYIV 3730
QY 402 -----D-----TG-----N-VT--R--TAA-H----- 412
Db 3731 LEKGPVPVPRPEVSADFVVISWEPRAYTGGCOISYIVEKDDTTTTHMMSATVAR 3790
QY 413 -----E-----D-A-IA-----VAS-AK-- 422
Db 3791 TTIKITLKTGTQYOFRIFAENRYGSAPLDSKAVIYQYFKEPQPGTPTTSISKQOM 3850
QY 423 -----N-----Q-----T-V--EF--D-----KV-- 431
Db 3851 LVQHEPVNDGCTKIIGHLEQKEKNSILMWKLNKTPIQDTKFKTTGLDEGLEVEFKYSA 3910
QY 432 -NI-G-GE--S--F-----K--Y-- 440
Db 3911 ENIGIGKPSKVSSECFVARDCDPPGRPEALVITRNNTLTKMKKPRAYDGSKITGYIV 3970
QY 441 -----R-----NI-----G-----A-F-- 446
Db 3971 KDLPDGRMKKASFTNVLETFEFTVSGLVEDQRYEFRIARNAAGNSESDSGAITARDE 4030
QY 447 -----F--D--K-----SK-----I 452
Db 4031 IDAPNASLDPYKQVIVVHAGETPVLLEADIRKPIPDVVMWSKDGKELETAARMEIKSTI 4090
QY 453 -TT-----N-----E-D-----G-T--K-- 460
Db 4091 QKTLVVKDCIRTDGQYIILKLSNVGTSKSIPIYKVLDRGPRPGLKVTGVTAEKYL 4150
QY 461 A--P-----S--K-----L--K-----F--V-- 468
Db 4151 AMNPLODGGANISHYIIEKRETSRLSWTQVSTEVOALNVKVTLLPGENEYIFRMAVANK 4210
QY 469 Y-IG-----K-G--Q-----D-----Q--D-----LI-- 479
Db 4211 YGIEPLESGPVTAQNPYKPPGPSTPEVSAITKDSMNVTVARPVDDGTEIEGYILEKR 4270
QY 480 -G-----L-DLR-----G--KIA----- 488
Db 4271 DKEGVRMTKCNKKTLLDLRLKLVGTGLTEGHSYEFVVAENNAAGVGPSEPSVFYRACALY 4330
QY 489 -----VMD-R-----IY-----TK--D----- 497
Db 4331 PRGPSPNPKVDTDRSSSVSLAMSKRIYDGAAPVKGVVEVEKAADDEVTCTPPTGLQGX 4390
QY 498 -----LK-NA--F-----KA-----MD----- 507
Db 4391 QFTVTKLKENTENYFRICAINSEGVEGEPATLPGSVAOERIEPPEIIDLARKVVLRA 4450
QY 508 -----KG-----A-----RA-I-----N----- 514
Db 4451 SATLRPLVTIKGRPEPEVKMEKABGILTDRAQIEVTSFTMLVLDNVTFRPSGRYNLTLE 4510
QY 515 -----Y-----VN----- 517
Db 4511 NNSGSKTAFVNVVRLDPSAPVNLITREVKDSVTLSEPPPLDGAKITNYIVEKRETT 4570
QY 518 -TV-----N-----Y--N-----R 524
Db 4571 RKAYATITNNCTKTTFRIENLOEGGSYFRVLASNEGYGLPAETTEPVKVEBPPLPGR 4630
QY 525 -D--N-----W-----T-- 528

```

```

Db 4631 VTLVDVTRNTATIKWEKPESDGSKITGYVEMQTKSEKSTCTQYVTKLATISGLTAG 4690
QY 529 -----ELP-----A-----MG- 534
Db 4691 EEYVFRVAAYNEKGRSDPRQLGVPIARDIEIKPSVELPFHTFNVKAREOLKIDVPFGR 4750
QY 535 -----YE-----A----- 537
Db 4751 PQATVMNRKDCQTLKETTRVNVSSKTVTSLSIENAKEDVGYELCVNSAGSITVPIT 4810
QY 538 -----DE-----G----- 540
Db 4811 IIVDRPGRPPIRIDEVSCDSITISMNPEYDGCQISNYIVEKETTSTTHIVSQAV 4870
QY 541 -T--K-----S-Q-----V--FS----- 547
Db 4871 ARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSPGPGTPKVVHATKS 4930
QY 548 -----I-----SG----- 550
Db 4931 TMLVTQVPVNDGSSRVIIGHLEKERSIILMSKANKILLADTOMKVSGLDEGLMEYRV 4990
QY 551 -----D-----DG-----V 554
Db 4991 YAENIAGIKSKSCSEVPVAPADPCDPPGOPEVTNITRKSVSLKMSKPHYDGAKITGYIV 5050
QY 555 -KL--N-----N----- 558
Db 5051 ERRELPGGRMLKCNVTNIQETFEVTELTEDQRYEFVARNADSVSEPESTGPIIVK 5110
QY 559 -----M-----IN-----P--D--K--TE-- 567
Db 5111 DDVEPRVMDVKFRDVIIVKAGVLKINADIAGRPLVISMWKGIEIEBRARTEIIST 5170
QY 568 -VK-----R--N-----N--K-- 573
Db 5171 DNHTLLTVKDCIRRDQGYVTLKNVAGTSVAANCYLDKPPRAGLEINGLTAERKS 5230
QY 574 -E--DFKD--K--L--E-Q-----Y-- 583
Db 5231 LSMGRPOEDGADL-DYIVIEKRETSRLSWTQVSTEVOALNVKVTLLPGENEYIFRVTGV 5289
QY 584 -Y--P--I--D-----M-----FS-----F-- 591
Db 5290 NKYGVGPELSVAIKALDPFTVPSPPTSLEITSYTKESMTLCSWRPESDGSSEISGYIIE 5349
QY 592 -NS-----N-KP--N--V-----G-D-E--K-----E--I--D- 605
Db 5350 RREKNSLRWVRVKKPYYDLRKVSTGLREGCEYERYVAENMAGLSLPSSETSPILRAEDP 5409
QY 606 -F-----K--F--AP-----DTD-K-- 614
Db 5410 VFLPSPSPSKPIVDSGKTTITIAWVKPLFDGABITGYVEYKKSDPTDMKTSIOSLRGT 5469
QY 615 E--L--Y-----Y-----KE-----DI-----IYPA 625
Db 5470 EYTTISGLTGAEVYFRVKSUNKGASDSSDPOIAKEREEPLFDISEMRKTLIVKA 5529
QY 626 G-S-T-S-----MG-P-----R--I--D----- 635
Db 5530 GASFTMTVPFRGRVVPVNLMSKPDJLRTARAYVDTTSRFSLTLENNRNRDNGSKYTLTIQ 5589
QY 636 -L--L-L-K--P-----DV--SA--P--G--KN-I-- 649
Db 5590 NVLSAASLTIVVKVLDPPGPPTNITVQDYTKESAVLSMDVENDGAPVKVYHIEKEBAS 5649
QY 650 -K--S--T--LN-----VIN--GKSY--G--Y 664
Db 5650 KKAWSVTNNCNRLSYKVTNLQEG-ALYFVRVSGENEF 5686

```

RESULT 14

```

AA
SQ Sequence 5701 AA;
Query Match 69.8%; Score 3070.2; DB 22; Length 5701;
Best Local Similarity 10.3%; Pred. NO. 7e-44;
Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495

QY 1 Y-----P-VV---LA-D-----T-----S-----9
      |         |         |         |         |
DB 73 YNFRISAINDAGVGEPVIPDVEIVEREMAPDFELDAELRRLTVVRAGLSIRIFVPIKGR 132
      |         |         |         |         |         |         |
QY 10 -----S-S-----E-----D-----A-----14
      |         |         |         |         |
DB 133 PAPEVTWKDNLNKNRANIENSTFTLLIIPECNRYDTGKFWMTIENPAGKSGFVNVR 192
      |         |         |         |         |         |         |
QY 15 -----LN-----I-----S-----D-----K-E-----21
      |         |         |         |         |         |         |
DB 193 VLDTPGPVNLNRPITDITKDSVTLHWDLPIDGSRITNYIVEKREATRKSYSTATTCKHK 252
      |         |         |         |         |         |         |
QY 22 --KV-----AEN-----K-E-----28
      |         |         |         |         |

```

QY 144 -----I-----E-T-T-----147
 Db 1393 SVSLSMWEKREHDGSGRLIGYIVEMQTKSGDKMCAATKATKTEATITGLIOGEEYSFRVA 1452
 QY 148 -----P---D-----N---L---D-K-----I-----154
 Db 1453 QNEKGISDPRQLSVPLAKDLVIPPAPKLTLEFTVLAGEDLKVDVPIGRPTAVTWHK 1512
 QY 155 -----KQ-----TE-----158
 Db 1513 DNVPLOKQTRVNAESTENNSLTTIKDAGREDVGHVVKLTNSAGEALETLNVILDKRGP 1572
 QY 159 --G-----I-----SS-----VE-----R---A-166
 Db 1573 PTGPNVKDEVTADSTILSMGPPKXDGSSINNYIEVKDSTTTMQIVSATVARTTIKAC 1632
 QY 167 -----Q-----K-----V-Q-----P---M---N-174
 Db 1633 RLKTGCEYQFRIAEENRYGKSTYLNSEPTVAQYPPKVGPPGTPVVTLSRDSMEVQWNE 1692
 QY 175 -----H-ARKE-----I-----GVE-----183
 Db 1693 PISDGSRVIGYHLERKERNLSILWVKLNKTPIPQTKFTTGLEBGEVEYEFVSAENIVGI 1752
 QY 184 -----EAI-----D-----Y-----L---189
 Db 1753 GKPSKVSECVYARDCDPGPRPEALIVTRNSVTLOMKKPTYDGSKITGYIVEKKELEPG 1812
 QY 190 --K-S-----I-NA-----194
 Db 1813 RMMKASFNTNIDTFEVTGLVEDHRYEPRVIAARNAGVFSEPSSTGAITARDEVDPRI 1872
 QY 195 --P-----F-----GK-----N-----F-----200
 Db 1873 SMDEKYKDTIIVHAGSEFKVDADIYGPPIPTIOWIKGDQELSNARLEIKSTDPATSLSV 1932
 QY 201 -----D-----G-----R---G-MVIS-----208
 Db 1933 KDAVRVDSGNITLAKAVAGERSVTNVVKVLDREPPGPPVVISGVTAEKCTLAMKPPPLQ 1992
 QY 209 -----N-I-----D-----T-----G---213
 Db 1993 DGSDDIINYIERETSRLVWTVVDANQVTLSCVTKLLEGNEVTFRIMAVNKKYGVGRL 2052
 QY 214 -----T-----D-----Y-----216
 Db 2053 ESEBVVAKNPFVVDAPAPAEVTTVTKDSMIVWVERPASDGSSEILGYVLEKRDKEGIRW 2112
 QY 217 -----R-----H-----KA-----220
 Db 2113 TRCHKRLIGELRLKVTGLIENHDYEFVSAENAGLSEPPSPSAVQKACDPIYKRGPPNN 2172
 QY 221 -----M-----221
 Db 2173 PKVIDITRSSVFLMSKRPYDGCCEIOGYIEVKCDVSGEMTMCPTPTGINKTINEVEKL 2232
 QY 222 -----R-----D-D-D-AK-----A-SMR-F233
 Db 2233 LEKHEVYFRICAINKAGVGEHADVPPIIVEKLEAPDIDLLELRKIIINIRAGSGLRLF 2292
 QY 234 -----K-K-E-----D-----L-----K-----GT--241
 Db 2293 VPIGRPTPEVVKWGVGEIDDAIIVTSSFTSLVDNNVRYSCKYTLTLENSSGTKS 2352
 QY 242 -----KX-----245
 Db 2353 AFVTVRVLDTPSPVNLKVTETITKDSVSIWEPPILLDGSKITKIYIVKREATRKSAAV 2412
 QY 246 -----W---L---S-----D-----KI-----251
 Db 2413 VTNCHKSMWKIDOLQEGCSYFRVTAENEYIGLPAQTAADPIKVAEVPQPPGKITVDVDT 2472

QY 252 -----P-H-----A-----F--255
 Db 2473 RNSVLSMTKREHDGSGKIIQYIVEMQAKHSEKMSSECAKRVKSLQAVITNLTOGEEYLFV 2532
 QY 256 --N-----Y-----Y-258
 Db 2533 VAVNEKGRSDPRSLAVPIVAKDLVIEBDVPKAFSSYSVOVGQDLKIEVPIGRPKPTITW 2592
 QY 259 -----N-----GGK-ITV-----E-----266
 Db 2593 TKOGLPLKQTRINVTDSLDTLISIKETHKDDGQYGITVAVNVGOKTASIEIVTLDKP 2652
 QY 267 -----KYDD-----G-----RD-----273
 Db 2653 DPKGPVKKFDVSAESITLSMNPPLTYGCGQITNYIVOKKDTTTVVDVVSATVARTTLK 2712
 QY 274 -----Y--F---DP-----277
 Db 2713 VTKLKTGEYQFRIAEENRYGQSFALSDPIVAQYPKBERGPPGTPPATAISKDSMVIOW 2772
 QY 278 H-----GMH-----I-----A-----GI-----LA-----287
 Db 2773 HEPVNGSGPVIGYHLERKERNLSILWTKVNKTTIHDQFKAQNLSEGIYEYEFVYAEINIV 2832
 QY 288 --G--N-----D--T-----E--Q-----D--I-----K---296
 Db 2833 GVCKASKNSECVYARDCDPGPTPEPIWVKRNEITLOMTKPVYDGSMTICYIVEKDLR 2892
 QY 297 -----N-----F-----N-G-I--D-G-I-A-305
 Db 2893 DGRMMKASFNTNIDTFEVTGLVEDQRYEPRVIAKNAAGAISKPSDSTGPITAKDEVLP 2952
 QY 306 --P-----NA-----308
 Db 2953 RISMDPKFRDTIIVNAAGETFRLEADVHGKPLPTIEWLRGDKLEESARCEIKNTDFKALL 3012
 QY 309 -----Q-I-----F-----311
 Db 3013 IYKAIIRIDGQYILRASNVAGSKSPVNVVKVLDREPPGPPVQVGTSEKSLTWSP 3072
 QY 312 -----S-Y-----K-----M---Y---316
 Db 3073 LQDGSDDISHVVEKRETSRLAWTVASEVNTLSKVTKLLEGNEVFRIMAVNKKYGVGE 3132
 QY 317 -----SDAG-----SG-322
 Db 3133 PLESAPVLMKNPFLPGRPKSLEVTNIAKDSMTVCNMRPDSGSEIIGYIVEKRDRSGI 3192
 QY 323 -----F-----AG-----DETMF-----330
 Db 3193 RWICNKRRIITDLRLKVTGLTEHDEYEFVSAENAGVGEPSPATVYTKACDPV-FKRG 3251
 QY 331 --H-----A-----I-----E--D-----S--I337
 Db 3252 PTNAHIVDTTKNSITTLAMKRPYDGCSEILGYVEICKABEEMQIVPTQGLKVTFRPEI 3311
 QY 338 -K--H-----N-----V-----D-----VV---S-345
 Db 3312 SKLTHEQBYKIRVCAKLVKVGISEATSVPGTVKPEDKLEBELDLDSRLKGIIVRAGGSA 3371
 QY 346 -----VS-S--G-FT-----G--T-----G-----L-----355
 Db 3372 RIHIPPGRPTPEITWSREGEFTDKQIEKGVNYTOLSIDNCORDNAGKIYILKLENSSG 3431
 QY 356 -----V-----G-----E-K---Y-WO-----A-----I-----R---365
 Db 3432 SKSAFVTIVKULDTPGPQONLAVKEVRKDSAFVLWEPPIIDOGAKVKNYVIDKRESTRKAY 3491
 QY 366 A-----L-----R-KA-----GTP-----MV---375
 Db 3492 ANVSSKCSKTSFKVENLTGALYYFRVMAENEFQVGVETVDAVKAABPPSPGKVTLT 3551
 QY 376 -V---AT-----G-----N-----Y-381

ID ABG23329 standard; Protein; 6619 AA.
 AC ABG23329;
 XX 18-FEB-2002 (first entry)
 DT Novel human diagnostic protein #23320.
 DE
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX MO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA
 PI Dmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 XX N-PSDB; AAS87516.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS
 PS Claim 20; SEQ ID No 53688; 103bp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue; as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WFO
 CC at fcp.wipo.int/pub/published_pcc_sequences.
 XX
 SQ Sequence 6619 AA;
 Query Match 69.5%; Score 3058.8; DB 22; Length 6619;
 Best Local Similarity 9.1%; Pred No.2.8e-43; Indels 5802; Gaps 489;
 Matches 586; Conservative 63; Mismatches 12;
 QY 1 Y-----P-----VV--LA-----D-----TS-----S----- 10
 Db 30 FITSPGKSGALVQIAEDFGKLIDHRLSTSADEDSGLPRFKENGRAEPAPDELFP 89
 QY 11 -----S-----E-----D-----AL-----N----- 16
 Db 90 LEGDAVPGFNGWLLFLDEFNSARKEIQAAAYKVLIDMTGQKQLHPYVAIVCAGNKATD 149
 QY 17 -I-----S-----D-----KEKVA-----E-----NK-----EK-H- 30

Db 150 RAITNNLSTAMQSRILHLEMETDFDVFMEVDVAIPNKMDERVIAPLANPNKLNDFEPHQ 209
 QY 31 ----- 30
 Db 210 EKTFCSTRTWEFNKIVSILPPGPINSEMTLLAGTITSGVATSFVQFOVYNNVSLNE 269
 QY 31 -----E--N-----I-----HS-----A--M-----ET--S 40
 Db 270 ILDNPKARMEEDNNLMAVVTSLINTDENHSHKIFDYERMPPTFKVLYYRSRKIDMS 329
 QY 41 -----D-----F--K-----E--K-----K-----TA----- 49
 Db 330 MTDANQLTREYDRQAKAFMGKTAAPFSSILCSLKFPMWVKDCGCDTAQDTGHELENNPDWE 389
 QY 50 -----VIK-E-----KEV-----V-----S-----KV-- 60
 Db 390 SLTPESRYTVMHLEIMVGLHSVRMGSRDPEVMNVACDIYINNQLIDGYFIGIENCM 449
 QY 61 -P-----V-----I-----D-----D-----NN----- 66
 Db 450 KDPYAGMAEEOIYDILMSKQIRPPKASGAFGTSGGDMKPSTSKASQVNIWNVVRAMH 509
 QY 67 ----- 69
 Db 510 QQKLSGSLPGVLPGRKMEVITQFLKRPVPMQVLLERPFNDLOETYYSWORPNRYPMY 569
 QY 70 -----E-----E-A----- 72
 Db 570 LPSPMDDGRLEHLAFLDTSGSIKPKDALFSSSEVAVYKEKQPOKMTLVQFTTEIYDE 629
 QY 73 -----K-----IK-----E-----E----- 77
 Db 630 QVIEDGKFTDVTIKGGGTSIVPVREWIHKPTAIIFTDLEVRMEPLRPDI PVIWV 689
 QY 78 --N--S-----N-K-S-----Q-----G-----D-- 85
 Db 690 VIRPAGQVPRGKGMVNSKSSGETTGLGAGVDIRIKQDVTVRRFLGLPCKIKINDEG 749
 QY 86 --Y-----T-----D-----S-F----- 90
 Db 750 KTKMKFGFTLASAMERFNVAGLVGIHDXSTWARTLSVFNVIIEPGMGVLTLELVY 809
 QY 91 -----VN--K-----N-----TEN----- 97
 Db 810 HGRKKLHPAGSGIAQVNGQVTVVARVAVTENSERGVITFTNGDSIDTOGYDSIR 869
 QY 98 -----PK--K--ED-----K-----VW--- 105
 Db 870 RNVAKALDGAQDIKERTPKLYQKTSBDGLANOKAIKAVSALVOGKODAVALLVVEQI 929
 QY 106 -----Y-----I----- 107
 Db 930 LVALLTAQVAKQKQETCTKATGALVWLYVDLADNFEMSDMKWQGISCVLAREGQELL 989
 QY 108 -----AEFK-----D-----KE-- 114
 Db 990 STLPKAEFKLAVTGAHKIVLGSIRTSLSYSRLSMINAKILADSITLSDGRVLVTMLV 1049
 QY 115 -----S-----G--EK--A----- 119
 Db 1050 YPRFIHSEFMTLRALNNRASSRAIPTORLIEOVRENEPWPVHWGKQKQASBELTTPM 1109
 QY 120 -IK--EL-----SS-----L-----K-----N----- 128
 Db 1110 EIKDAEFIVENVAASAAVVAEQLRGQVNHQIVRILIEPTFHIRVVVTATQWNNPFGRLI 1169
 QY 129 -----TK-----VL-----Y-T-----YD-----RI----- 138
 Db 1170 HPDAQPEICELATYMKEOYDSKPRTLTLAGEMHLPIYITNSDYLDAYDYCKKHQIRTRDEPS 1229
 QY 139 -----FNG--SAIE-----T-----TP-- 148

Db 1230 SDEINCLLVKISAAACARASNNFKGTPSTIEDLRLFSKLVSFVHASPTEHQATPOH 1289
Qy 149 -----D-----N-----LD-----K--IK-----Q----- 156
Db 1290 VMSKGLNEPTDPTWENGISHMDRNLRLYSGGHMAKPRFKRIKDKPVMGSAAGQSHKGD 1349
Qy 157 I--E-----GI-----S-----S-----V-----E--RA-----OK----- 168
Db 1350 IYVECLQYDGVHAHDTRATELUSKERQLWLSPNKMTWIOADLAEVQRMILDEWSEDOFIDG 1409
Qy 169 -----V-Q-----P-----N-----M-----K--E----- 172
Db 1410 NAEAITDDMASDMVFLQAMAEKPDSTELTTTSATLGAQADMAFGISEDPAFFQILSANL 1469
Qy 173 -----M-----N-----HA-----R-----K--E----- 179
Db 1470 YSNQKLAMVRETLCSNWDHAGNKTDIPIHTIDKDGFLIFRDYGSIGPKDKIOEVYGV 1529
Qy 180 -----I-----N-----G----- 181
Db 1530 YGASTKXANTAGTGGFLGCKSPMAYTDSFTVSMCDGKTGYHYHIAKSSVETNGKPGIIP 1589
Qy 182 --V--EEA-----ID-Y--L-----KSI-----NA-----P----- 195
Db 1590 IVSVPTESGLVKIQLQSEDIIDSYIRAILVHGEIKCIYCNQYTGNTVELLPVLGM 1649
Qy 196 -F--G-----K-----N-----F----- 200
Db 1650 SPEGSYNLDSHTWKHYMGSHSIFIRYGNVMYPALDTPATHEALKLVKQFMQVINCRI 1709
Qy 201 -----DG-----R-----GM----- 205
Db 1710 VVQAAPSSLAVAPSRETLSSQKMTEDGIVNLAVLVDMKADIKARLPNAIEYIGKAIKE 1769
Qy 206 -----VI-----S-----N-ID----- 211
Db 1770 YSPHRFEVNSYLPYTNLPDVIIVSRYSMSSTLSKQYAHCKKGNWEIDRQLEHAPALAHF 1829
Qy 212 TC-----TD-----Y----- 216
Db 1830 SGSQKAHLKHLKRAKDLNRYRELTKVFTDMAIKPMFKTFAKLGITKGISLFHYQYGVEL 1889
Qy 217 -----R-H----- 218
Db 1890 WRNNIVNSVRTHSLNFMFTYSSKKCRVFTVPRITNIEDSLAGYEPEDNLSTETLVKYNIS 1949
Qy 219 -----KA-----M----- 221
Db 1950 IVKVGTKKEAEAMTNLLTKAGYEVINMAEYHEWDCIARRRKEEQEAKLAISKLOAKKAT 2009
Qy 222 -----R-----I-----D-D-----D----- 226
Db 2010 NKKVSPNRLISWNILEIPKEHRSYSSEYCGSHYASRSHDRMVDVDTPOYIMQEDINNY 2069
Qy 227 -----A-KA-S-----MR-----F-K 234
Db 2070 TGRGLVSMQLTPATIKAVTVVCRNKIERNAIKRGAVHLDWTAPYIASVLTSSQFIK 2129
Qy 235 --KE-----D-----LKG----- 240
Db 2130 YVTKERQKTLSDIGIYSKHLRLRLDITYAPLKLGLRYDPELENAVLFLOSSYPKLRAL 2189
Qy 241 -----T-----D-K-----N-YW-----L 247
Db 2190 LGKGLITEEYHFLSDVDVHCKTFNSRLTLQPDGSTMVIFOGDARVPRIVAEAKEKNL 2249
Qy 248 S-----D-----K-I-----PH----- 253
Db 2250 TPGKIIEDVDTIEAQIRQBFDAERDTGGFVKFRITTKKLOEFPLGEBEEQVPHVSPVE 2309
Qy 254 -----AFN-----Y-Y-----N----- 259
Db 2310 LGAKPTAFNAFANAVGKTLAVQHASLKAPEMTVTHERFYGNVAVMTDVPDRNHRIE 2369

Qy 260 -----G-----G-----K----- 262
Db 2370 VIHVIRESGTSEYVNNLLAQPTPFKVFVDWLTETQAKFVRRLNNTOGVISVYNKLEBDG 2429
Qy 263 I-----T-----V-----E----- 266
Db 2430 IPVKVAVKATEPTNKEKLEAASTKLKANGAVATDAYEFTDLNENEETIIVAVVNDKVVDP 2489
Qy 267 -----K-Y-----DG----- 271
Db 2490 VQNLHRHIQSANLKDYGFTKFLERLSTVIDKRRHSVEDELMKFMKGDLPLIADDSGIVI 2549
Qy 272 -----R-D----- 273
Db 2550 FKRLSKDNPEHLKDVMDVFDCHSKKIEOCVGMVMVKNLVQNRKRCDCSHGLHVAS 2609
Qy 274 --Y-----F-----D-----P-----H----- 278
Db 2610 LOYIRNFSGNVTIICKVAPEDVFVPEYDVTKMRVSAVHIITKLPDAIRDHVNNGNPIST 2669
Qy 279 --G-----M-----H----- 281
Db 2670 IEGGTEILNMVLSGNHPSRQVLVGGHYGNLTYSSELTRNYLDDDESVIPDEPVIKKT 2729
Qy 282 -----TA----- 283
Db 2730 ALNMEESFTATNPKAEPVKATDVKPIKTKMEILKDLWNQFLKAETSNEAITLADEIIAA 2789
Qy 284 -----G-----I----- 285
Db 2790 KGKCKKSNASLGFSQDMVQKLIIDARANKPKEKAVQDKTVAPKTTVTRSRNADVIRSY 2849
Qy 286 L-----A-G----- 288
Db 2850 LNDKMSDYCKAHAIDHVKRAAKSYAALGLTPBECSDIKLKHHLKIHTWAKVYRANK 2909
Qy 289 -----N-----DT----- 291
Db 2910 AHDDIIRNNAVGLSGTIAKTGLGVHPTVTLRSLGIEPADTRRTFMENVLAPLPHV 2969
Qy 292 -----EQ-----D----- 294
Db 2970 ADMLSEQVGPKEIRQYVYRDLLLLEAYNNRHHKETAHERFLKRYVVEVMELELAPNSEE 3029
Qy 295 -----I-----K----- 296
Db 3030 AYDTLHNAIAALTELSELMSKSSDKVYAPNSNRLAILDALADQIVTATGIGTFLGMNVPG 3089
Qy 297 -----N-----FN-----G----- 300
Db 3090 ALTEVNRNYSKPEDGEPVFNENKVMKEAAAEAGFFDLDPNASELRISGPGTGKTF 3149
Qy 301 -----ID-----GIAP-----N-----A-----Q----- 309
Db 3150 MAHMIDEIMPRYHETCSLMGI-PALYNEVIMTATTNKAAEVLARATGRPTSTYHSFQGLT 3208
Qy 310 -----IF-----S-----Y-----K----- 314
Db 3209 VRNLLKTEADLIPSKSFISIKKNKVFIDEASMDRKLKPFINEGTHOSKIVYLGDKCOL 3268
Qy 315 --M-----Y--S-----D-----A-----GS-----GF-----A----- 324
Db 3269 LPVKETSSPVNDSSIKSFMLTQQMRTDVPPELHALHEQLRGIEGKHGFLPIKAVPGIID 3328
Qy 325 -----GDE-----T-----M--PH-----A 332
Db 3329 WTKGDEMKEVEGHFTKTDTRIIVAYTNOQVYNTFIRAMNGFHGEFICEELVNSA 3388
Qy 333 I-----E-----DS----- 336
Db 3389 IQIGAGDRLSIBQEQIVDQDSATRKILIDSSGIELEVRDCTLDTCYGVGFEGIPVDM 3448

QY 337 --I-K--H--N-----V----- 341
Db 3449 DYFIRLQGYVYAHQKWEHFFYLKETFPBLRALHASTVHKSGSGSTVDTFIDATDLSTGRQ 3508
QY 342 -D-|-----VS-----V-----SS----- 348
Db 3509 PDVAVRLLVYAVSRARHRVVEYGSIAEKFGGELMQLGKQIRREMTNSSELVOELFLAELR 3568
QY 349 -----G--F-----TG-----TG----- 354
Db 3569 RLDKVLDSTITRNGETFAVYNSTGFLYLGEFYSSRSQOPSPGHRLLALANHLDDMKREY 3628
QY 355 -----LV-----G----- 357
Db 3629 LEQAGKLLLEVHMVNOVFRLYQGCKTYQDVADALPECLIAMDKGHLKSQRTPKAFT 3688
QY 358 -----E-----K- 359
Db 3689 LDNDAMALRHNRKSGRLSNTWREPILSFEVCMRYITFPNNEGIVPAILTLTLRKDDIVKA 3748
QY 360 Y--W-----O-----A-----I----- 364
Db 3749 YMEFMDVDKDSVMIELHTAPGKKKTPAKEMOEFTQELVPAFEAAQTQYIICGDSERYK 3808
QY 365 --R-----A-----LR-----K-----A----- 370
Db 3809 QLTQKQKADANIGVYMDCAVGNQKVIYVPRNQIFYPDPDKIKTKIALSMQALLIDVSGSY 3868
QY 371 --G--I-----P-----M-----V-VAT-----G-----N- 380
Db 3869 VEPGKTIHVADYPLTPRAISKWLDKLEMDVPLAVDIETFSLKHYDGGIGITFCWNKH 3928
QY 381 -----Y-----ATSA-----S----- 386
Db 3929 EGIAFPVDYEPREGATEAPYRGQVHNMLVRSMLRDFIKYLNRMQYHNIAFDVYALIYQL 3968
QY 387 --S--S-----S--WD--LV-----A-----N 395
Db 3969 FMTLLDPTGGLHGHGSIMLRWMDCTKLITYLATNSCAKNKSLKDQAOEYAGNAAQOEIN 4048
QY 396 -----N-----H--L--KN--T-----D-----TG- 404
Db 4049 DITRIPLAELLEVLVDGLCTWYVEKHMDTLVNDQDLVYTNIFKPACEDIIQWQLTGM 4108
QY 405 -----N-----V-----TR----- 408
Db 4109 PINNRVLEVEEALTDYNNALKTITADSKVIKDFTRLNEEWEKONQILKKKRYTLADA 4168
QY 409 -----TA--A-----H-----E--DA-I 416
Db 4169 KEQFNPSGIOLQKLLFEFLGLPVGLTASKLPRATGSGILSKLKHOTDTSILEILDALI 4228
QY 417 --AV-----A--GA-KN--Q-----TV----- 426
Db 4229 DYKAVDKLITFAIFALKNARQGPDMHYLFGNLNLGTVSGRLSSSEBNQLNPSGRYA 4288
QY 427 -----E-----F-----D--K----- 430
Db 4289 KMIKCEAPRGWIFCGLDFASLEDRISALTTPKPNKLKYVTDGDGSHLRAYVFGEM 4348
QY 431 -----VN-I-----G-----G-----E----- 436
Db 4349 PDIEDVESVNSIQEKYAYRODSKAPTFALTQCTYITLMKNGCFPEQAKRMBEERYHT 4408
QY 437 --S-----F-----K--Y----- 440
Db 4409 LYKFSDDVQAQLDQAAKQGVTVAFGLRVTRPLQVIRGTSKTPYBAEAGRTAGNAL 4468
QY 441 -----RN----- 442
Db 4469 GQSMCLLNRAQSEFMKRVKNSKRLDIRPSTHIDAQYFLIRDDMDVVIYTNHLVKAV 4528
QY 443 -----I-----GA----- 445

Db 4529 QWQDPRIDAPDVLHDQGEAGVIVQNTMLLTBEKQVTFRDLGRAQSALFQNLNDRFGQV 4588
QY 446 -----F-----FD----- 448
Db 4589 DVRDIVFLNYSYLGMSQRPDQONLVNPPORMEOPRYOAECEDIIOALNMTFDSGLFKA 4648
QY 449 -----K-----S--K--I-- 452
Db 4649 LMRVAAARLGNGKRGKGNNAVYDAEKVHYANRIFIKESGGQKKEABEPVSAALKKGEEIHM 4708
QY 453 -----T-----T--N-----E-----DGT----- 459
Db 4709 SLSELFTLLTGEAAEGEWSLAGMHTVPRNLGVMPRIAVMSKAEQDVVADGTTIHHV 4768
QY 460 -----K-AP----- 462
Db 4769 FPHBIDWSDKIAFPARITAMRFVWLHDEYDVVKDKKYSVTLTLPKIKHVMKHVLD 4828
QY 463 -----S-----KL----- 465
Db 4829 SVESMDIMDVSTSMGTGLHDSIEKAMNNGHNALRKLKCYPERVQVNVINPTREQIASN 4888
QY 466 -----KF-----V-----Y--I--GKG-----Q-- 474
Db 4889 PNLIPWIEQATKKIAGWTIGKFDIVTBSGLDQVNSTYTTWVGK-GRDEEHRLOGSM 4947
QY 475 -----D-----O-----D----- 477
Db 4948 YRWLHDDKITEDVIRINVIPTDMKKALATSGENYRKRVHLKDIPLMSYEATEEMIKAKL 5007
QY 478 -LI--G-----L-----D-----L--R----- 484
Db 5008 ALIDKYWDAPSEIPECTDEBELMRSDVFKYFAPDEKAKOGARSTKNPFTLFEARKFMA 5067
QY 485 --GK-----I-----A-----VM-----DR-----I-Y----- 494
Db 5068 EKGKGTILHVEIFRPMJDLTGVSHNHRAIEIVDLCKTQONTDGFRRVAVAYFLAKMAS 5127
QY 495 -----TKD-----L-----K--N-----A--FKK--AM----- 506
Db 5128 CMGATITTKORGEIIPVNIYAVVALATSGFGKHSVNIVEDGFMAGFKKGFMEDTPALAE 5187
QY 507 -----DK-----GA-----R-----A 512
Db 5188 RLWKIANERSARNAATDOKEEFDKVEAEYRTGAYPTFDSGTPRAVQOLRHKLLAAGCA 5247
QY 513 I--I-----M--V--V-----NT-----V----- 519
Db 5248 INLOIDEIGSNLANADVLTVLELYDOGKIKOKULTKNTAESIRSEEDGKTPANLLFG 5307
QY 520 -----N-----Y-----Y--N--BDN----- 526
Db 5308 TPSKILNGSQTEBELFYDLDTGYSRCLFAVGAQDKBRAVLSQSAEITYRNLIKODNNAV 5367
QY 527 -----W----- 527
Db 5368 NKWANHFSLADASKFGMKQTVEDDVGIALIDYKIOCEKQAAALADHEIRKAELEHRYF 5427
QY 528 ----- 529
Db 5428 KALYLAQAFAPVDSINVEHENLKQALLVEESGVAFOTILNREKAYVLAKYIASVTE 5487
QY 530 -----LP-----A--M-----GY----- 535
Db 5488 VTHADLEALPFPKSGNAARNEWMTLATWGYKOHIIIKKTYNGIEFFRGETLKETDTN 5547
QY 536 -----E-----A-----D-----E----- 539
Db 5548 EMTVSGENPAYDYIAEKVPDQHLVLTQOAGYHMANHFRGRHAEENALAGFNMIVID 5607
QY 540 --G-----TK--S-----Q-----V-- 545

Db 5608 VDGTCSLDVCHELMKEYRPMFTYTTKHTDEENRFRLLIPNNYQLELDGAEBYKPFMDNVMA 5667
QY 546 ---F---SI---S---G---D---D---D---V---V--- 554
Db 5668 WLPFKTDESANQRAKKWESCAKGTYYYNDAPLLDVRDFIPRTSKNEAFQOGMKVQNLD 5727
QY 555 KL--W-----N-MI-----NP-----D--- 563
Db 5728 NLERWFAQRIATGNRNHMIKALALVDSGIGFNEVKQRVHEFNKKLSNPLSADLSTV 5787
QY 564 ---K---K---T---E---V---K---RN----- 571
Db 5788 MITVAKLLFLVEEDCQLKEETNSEEFHGEWNDOLILIVGFSAAKGSASLRNINQERWV 5847
QY 572 --N---K-----E-----D---F-KDKLE-QY 583
Db 5848 YLNTAEAGKRLPFKNFNGVRITDPYQILEYFDQIEHRDDVDGIIIVDSLTFMMDMLETQY 5907
QY 584 -----Y-----P-I-----D---ME-SF----- 591
Db 5908 VLTAAANTQKAWGEFAOFFKILQEKVVKFAKPVIFTAHVKDEVDERAMELKTFFVPVKGSL 5967
QY 592 --N-----SN-----K-----P---NVG----- 599
Db 5968 KNGIEAYFTSVSAERIDLKELEKYSNGMLEITEDEQELGYKHVQFTRTKTVGKRIR 6027
QY 600 -----D---E---K-E-ID---F----- 606
Db 6028 SPNGMFSKQTYMNDNAQKLLDLHAEYYSTGQAAKAEKATDNLGGFGAKESDIYLANLK 6087
QY 607 ---K---F---APD-TD-----KE-LY-----KE----- 619
Db 6088 VAYAGKAASGANFIQIIA-DLTDLDHGSAGEYREQLYITSGTEKCKCTYEKNGKEYFLP 6146
QY 620 ---DI---I-----Vp--A--G-----S 627
Db 6147 GYTVINDILVMTSGETIPEAVFEKVVNVYDFDEKKEVAKSVMPVNAIGKFAVALKS 6206
QY 628 ---T-----S-----K----- 629
Db 6207 EEDKQTKDGSNGVYSTGETRFTNTIEKVPHDLHLTVVEABELTERGKELTVEBAVFWDK 6266
QY 630 W-----G---P----- 632
Db 6267 WLEKNKGVTDRDKTTKGASGAKGQPPKPGATNTGAGASAAKSLFEIMOIPVLGVDPSPFN 6326
QY 633 ---R---I---DL-LL--KPD---V--S-----A--- 644
Db 6327 WGLARGMLDLETGILSLGLDLKLVTGDKQVRQNSKDMQAAEDITTVIDWFKEAKVI 6386
QY 645 ---P-G---KN-IKS-----T-LN---V--I-N-----G-K-S----- 660
Db 6387 FVEVPVGSQSANGMSKYGVCVIGLALTALGHEIIEVTPIENKVALSGIKTASKDMVIRA 6446
QY 661 -T--Y-----GY 664
Db 6447 ATGFYPEANWLKDTKEFNENFIKGITMQVILTQPEIEAALTGY 6489